IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 6, July 2010
ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

A New Approach to Supervise Security in Social Network through Quantum Cryptography and Non-Linear Dimension Reduction Techniques

Lokesh Jain[1] and Prof. Satbir Jain[2]

[1] M.Tech. Final Year, NSIT,
Dwarka, New Delhi, INDIA

[2] Department of Computer Science, NSIT,
Dwarka, New Delhi, INDIA

Abstract

Social networking sites such as Orkut, Tribe, or Facebook allow millions of individuals to create online profiles and share personal information with vast networks of friends - and, often, unknown numbers of strangers. Some of the information revealed inside these networks is private and it is possible that corporations could use learning algorithms on the released data to predict undisclosed private information. To find the patterns of information revelation and their security implications, we analyze the online behavior of wiki-vote data set and evaluate the amount of information they disclose and study their dimension used for reduction. In this paper we conclude that dimension reduction is one of the factors through we can achieve the security and maintain the integrity of dataset. We highlight various Non-Linear dimension reduction techniques with quantum cryptography to produce the desire result and show the comparative result with linear dimension reduction technique.

Keywords: *Social network, Security, Quantum cryptography, Non-Linear Dimension reduction*

1. Introduction

Recently, online social network has emerged as a promising area with many products and a huge number of users. With the development of information retrieval and search engine techniques, it becomes very convenient to extract users' personal information that is readily available in various social networks. Malicious or curious users take advantage of these techniques to collect others' private information. Therefore, it is critical to enable users to control their information disclosure and effectively maintain security over online social networks.

One of the challenges in social network is security. Although security preservation in data publishing has been studied extensively and several important models such as *k*-anonymity and *l*-diversity as well as many efficient algorithms have been proposed, most of the existing studies can deal with relational data only. Those methods cannot be applied to social network data straightforwardly. Security may be break if a social network is released improperly to public. In practice, we need a systematic method to anonymize social network data before it is released. However, anonymizing social network data is much more challenging than anonymizing relational data on which most of the previous work focuses.

One of the ways to achieve the security in Social network is to reduce the dimension of whole social network by linear and non-linear dimension reduction technique. The dimension of the data, is the number of variables that are measured on each observation.the problems with high-dimensional datasets is that, in many cases, not all the measured variables are "important" for understanding the underlying phenomena of interest. While certain computationally expensive novel methods can construct predictive models with high accuracy from high-dimensional data, it is still of interest in many applications to reduce the dimension of the original data prior to any modeling of the data. By reduce the dimension of these models we can also achieve the security of the original data set.

So in this paper, in the second section we introduce the social network analysis. In the third and fourth section we refers to the non-linear dimension reduction techniques and in next two section we represents our design approach and experimental result and in the last section we concluded with future work and conclusion.

2. Social Network Analysis

Social network analysis [SNA] is the mapping and measuring of relationships and flows between people, groups, organizations, computers, URLs, and other connected information/knowledge entities. The nodes in the network are the people and groups while the links show relationships or flows between the nodes (Fig.1). SNA provides both a visual and a mathematical analysis of human relationships. Management consultants use this methodology with their business clients and call it Organizational Network Analysis [ONA]. To understand networks and their participants, we evaluate the *location of actors in the network.* Measuring the network location is finding the *centrality* of a node. These measures give us insight into the various roles and groupings in a

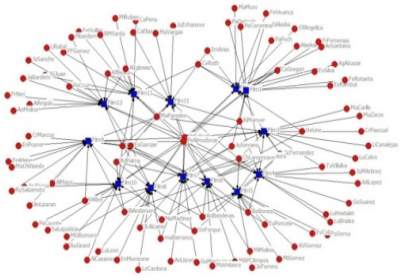


Figure 1: social network structure

network - who are the connectors, mavens, leaders, bridges, isolates, where are the clusters and who is in them, who is in the core of the network, and who is on the periphery?

So in nut shell we can say that Network analysis is the study of social relations among a set of actors. It is a field of study -- a set of phenomena or data which we seek to understand.

In the process of working in this field, network researchers have developed a set of distinctive theoretical perspectives as well. Some of the hallmarks of these perspectives are:

- Focus on relationships between actors rather than attributes of actors.
- Sense of interdependence: a molecular rather atomistic view.
- Structure affects substantive outcomes.
- Emergent effects.

Network theory is sympathetic with systems theory and complexity theory. A social network is also characterized by a distinctive methodology encompassing techniques for collecting data, statistical analysis, visual representation, etc.

3. Reason for choosing Non-Linear Dimension Reduction Technique

In case of social network, the size of the data set is large and data set has various dimensions. Due to severity of social network data, it is very difficult to secure the data. Consider a dataset represented as a matrix (or a database table), such that each row represents a set of attributes (or features or dimensions) that describe a particular instance of something. If the number of attributes is large, then the space of unique possible rows is exponentially large. Thus, the larger the dimensionality, the more difficult it becomes to sample the space. This causes many problems. Algorithms that operate on high-dimensional data tend to have a very high time complexity. Many machine learning algorithms, for example, struggle with high-dimensional data. This has become known as the curse of dimensionality. Reducing data into fewer dimensions often makes analysis algorithms more efficient, and can help machine learning algorithms make more accurate predictions.

By reducing the dimension of dataset, we can also achieve the security. Because dimension reduction only represents the abstract feature of a particular data set. Data abstraction shows only those features that are essential to represent the data and hide the remaining details. Hence data hiding is a one way to achieve the security. In this paper we analyze that non-linear dimension reduction a more efficient way as compared to linear Dimension reduction techniques. These techniques not only use for feature selection and extraction but can also be used for security purposes.

4. Introduction to Non-Linear Dimension Reduction Techniques

Advances in data collection and storage capabilities during the past decades have led to an information overload in most sciences. Researchers working in domains as diverse as engineering, astronomy, biology, remote sensing, economics, and consumer transactions, face larger and larger observations and simulations on a daily

basis. Such datasets, in contrast with smaller, more traditional datasets that have been studied extensively in the past, present new challenges in data analysis. Traditional statistical methods break down partly because of the increase in the number of observations, but mostly because of the increase in the number of variables associated with each

observation. The dimension of the data is the number of variables that are measured on each observation.
We subdivide techniques for dimensionality reduction into convex and non-convex techniques (Fig. 2). Convex techniques optimize an objective function that does not contain any local optima, whereas non-convex techniques optimize objective functions that do contain local optima. The further subdivisions in the taxonomy are discussed in the review in the following two sections:

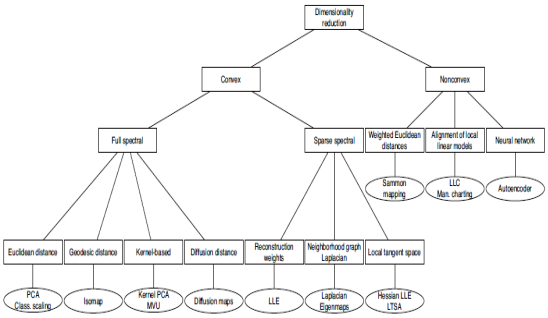


Figure 2: Taxonomy of Dim. Reduction techniques

4.1 Convex Techniques for Dimensionality Reduction

Convex techniques for dimensionality reduction optimize an objective function that does not contain any local optima, i.e., the solution space is convex. Most of the selected dimensionality reduction techniques fall in the class of convex techniques. In these techniques, the objective function usually has the form of a (generalized) Rayleigh quotient: the objective function is of the form $\emptyset(\mathbf{Y})=\frac{\mathbf{Y}^T\mathbf{A}\mathbf{Y}}{\mathbf{Y}^T\mathbf{B}\mathbf{Y}}$. It is well known that a function of this form can be optimized by solving a generalized eigenproblem. One technique (Maximum Variance Unfolding) solves an additional semidefinite program using an interior point method. We subdivide convex dimensionality reduction techniques into techniques that perform an eigen decomposition of a full matrix and those that perform an eigendecomposition of a sparse matrix.

4.1.1 Full Spectral Techniques

Full spectral techniques for dimensionality reduction perform an eigendecomposition of a full matrix that captures the covariances between dimensions or the pairwise similarities between datapoints (possibly in a feature space that is constructed by means of a kernel function). In this subsection, we have five such techniques: (1) PCA / classical scaling, (2) Isomap, (3) Kernel PCA, (4) Maximum Variance Unfolding, and (5) diffusion maps.

4.1.2 Sparse Spectral Techniques

In the previous subsection, we discussed five techniques that construct a low-dimensional representation of the high-dimensional data by performing an eigendecomposition of a full matrix. In contrast, the four techniques discussed in this subsection solve a sparse (generalized) eigenproblem. All presented sparse spectral techniques only focus on retaining local structure of the data. We discuss the sparse spectral dimensionality reduction techniques (1) LLE, (2) Laplacian Eigenmaps, (3) Hessian LLE, and (4) LTSA

4.2 Non-convex Techniques for Dimensionality Reduction

We have a non-convex techniques for multidimensional scaling that forms an alternative to classical scaling called Sammon mapping , a technique based on training multilayer neural networks ,and two techniques that construct a mixture of local linear models and perform a global alignment of these linear models. So we have following non-convex techniques (1) Sammon Mapping, (2) MultilayerAutoencoder, (3) Locally Linear Coordination (LLC) and (4) Manifold Charting.

5. Introduction to Quantum cryptography

Quantum cryptography was proposed by Bennett and Brassard in 1984, who also defined the first QKD protocol, called BB84. At time of writing, a handful of research teams around the world have succeeded in building and operating quantum cryptographic devices. Fundamental aspects of quantum physics – unitarity, the uncertainty principle, and the Einstein-Podolsky-Rosen iolation of Bell's inequalities – suggest a new paradigm for key distribution: quantum cryptography. Initial experiments seem to confirm the utility of this paradigm. Assuming that the theoretical models continue to be confirmed in the use of actual devices, the fundamental laws of nature can be invoked to assure the confidentiality of transmitted data. Quantum cryptography – more properly termed Quantum Key Distribution, QKD – employs two distinct channels. One is used for transmission of quantum key material by very dim (single photon) light pulses. The other, public channel carries all message traffic, including the cryptographic protocols, encrypted user traffic,

etc.QKD consists of the transmission of raw key material, e.g., as dim pulses of light from Alice to Bob, via the quantum channel, plus processing of this raw material to derive the actual keys (Fig. 3). This processing involves public communication (key agreement protocols) between Alice and Bob, conducted in the public channel, along with specialized QKD algorithms. The resulting keys can then be used for cryptographic purposes, e.g., to protect user traffic. By the laws of quantum physics, any eavesdropper (Eve) that snoops on the quantum channel will cause a measurable disturbance to the flow of single photons. Alice and Bob can detect this, take appropriate steps in response, and hence foil Eve's attempt at eavesdropping.

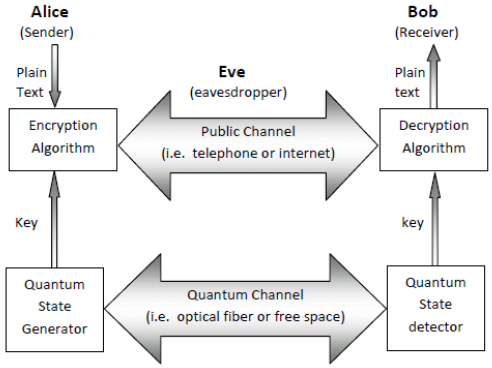


Figure 3: Quantum - Key Distribution

6. Design Approach

For achieve security in social network, we proposed a model using quantum cryptography to achieve security (Fig. 4). In this model first we collect the information regarding the social network data and after that reduce the dimension of the data set using various dimension reduction technique. After reducing the data, we covert the data into digitized form and to encrypt the data we use the MD5 (Message Digest 5) technique. Generally, instead of MD5, we can also use any other encryption techniques, such as DSA, AES and so on, to decrypt the message. For the secure key distribution purpose we can apply the quantum cryptography. Quantum cryptography provides a secure way to distribute the key on the basis of quantum theory. Generally, it is very difficult to implement this key distribution technique because it requires Photon light pulses (PLP) through which photon travel from one user side to another side. So, in this paper we just proposed the QKD (Quantum Key Distribution) for key distribution. After encrypt the key we transfer the data through SNSs site and at the other end, user decrypt the data in the reverse order.

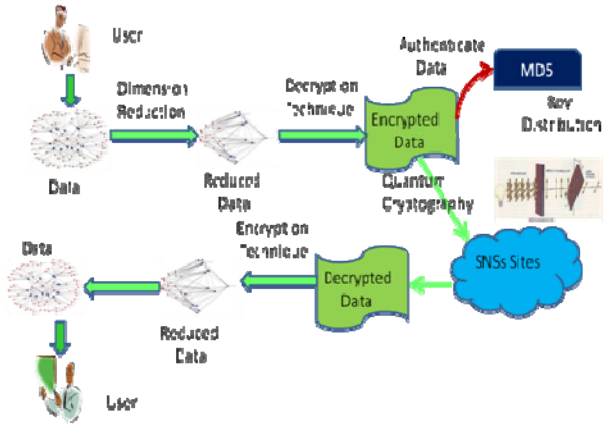


Figure 4: Design Approach to preserve security

Hence, in this paper we proposed a secure and authenticate technique for social network data sets which is more secure as compared to earlier techniques.

7. Experimental Result

In our experiments on 'wiki vote' datasets, we apply the ten techniques for dimensionality reduction on the high-dimensional representation of the data. Subsequently, we assess the quality of the resulting low-dimensional data representations by evaluating to what extent the local structure of the data is retained. The evaluation is performed by measuring the generalization errors of 1-nearest neighbor classifiers that are trained on the low-dimensional data representation.

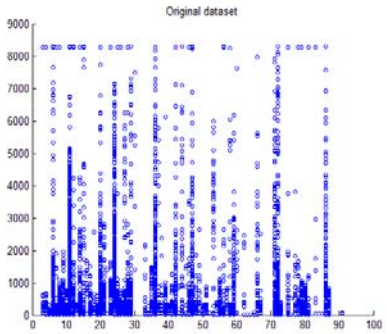
Wiki-is a free encyclopedia written collaboratively by volunteers around the world. A small part of Wikipedia contributors are administrators, who are users with access to additional technical features that aid in maintenance. In order for a user to become an administrator a Request for adminship (RfA) is issued and the Wikipedia community via a public discussion or a vote decides who to promote to adminship. Using the latest complete dump of Wikipedia page edit history (from January 3 2008) we extracted all administrator elections and vote history data. This gave us 2,794 elections with 103,663 total votes and 7,066 users participating in the elections (either casting a vote or being voted on). Out of these 1,235 elections resulted in a successful promotion, while 1,559 elections did not result in the promotion. About half of the votes in the dataset are by existing admins, while the other half comes from ordinary Wikipedia users (Table 1).

The network contains all the users and discussion from the inception of Wikipedia till January 2008. Nodes in the network represent wikipedia users and a directed edge from node *i* to node *j* represent that user *i* voted on user *j*.

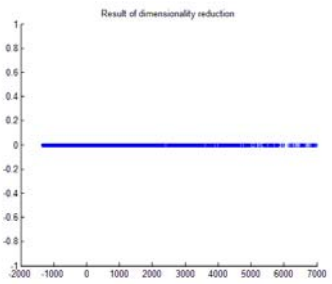
Table 1: Representation of Wiki-Vote Data set

Nodes	7115
Edges	103689
Nodes in largest WCC	7066 (0.993)
Edges in largest WCC	103663 (1.000)
Nodes in largest SCC	1300 (0.183)
Edges in largest SCC	39456 (0.381)
Average clustering coefficient	0.2089
Number of triangles	608389
Fraction of closed triangles	0.1255
Diameter (longest shortest path)	7
90-percentile effective diameter	3.8

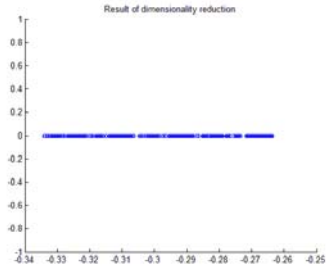
We perform the Non-linear dimension reduction techniques on the wiki-vote data set and get the following result:



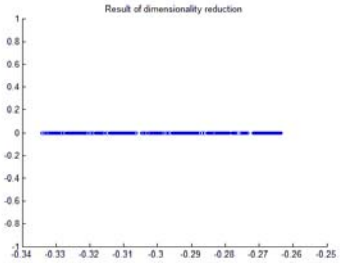
Original data set



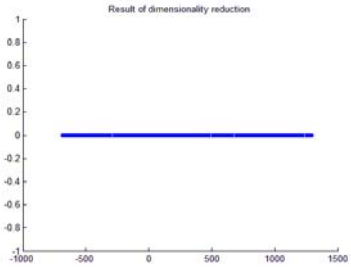
PCA



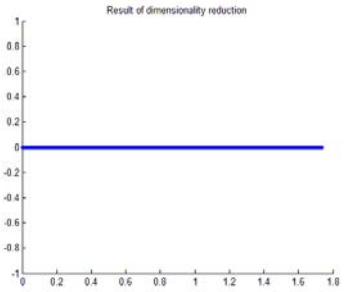
Isomap



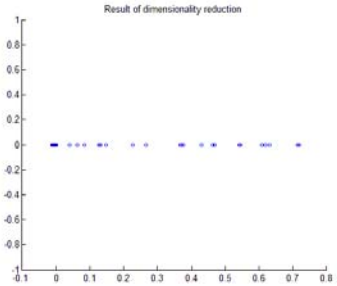
Autoencoder



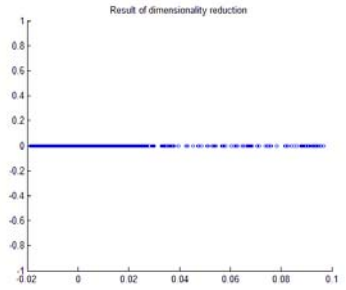
Diffusion Map



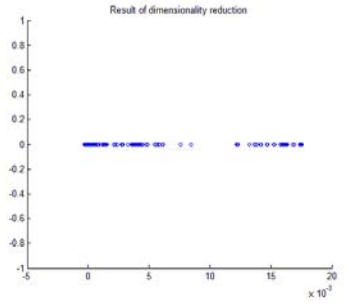
Hessian LLE



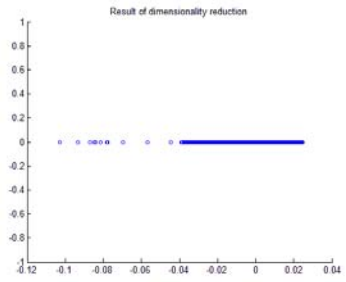
Kernal PCA



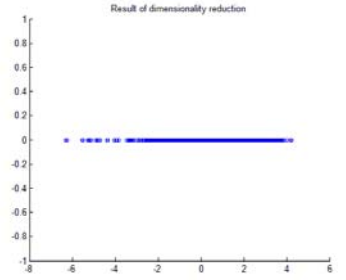
Sammon Mapping



Laplacian Eiganmap



LLE



LLC Man. charting

In the above diagram we represent the graphical representation of the dimension reduction of various Non-Linear techniques.

After reduce the dimension of data, we convert the reduce data into digitized form by sampling technique and encrypt the data using MD5 technique. Hence, we can find a way to achieve security in social network.

Now, we represent the computational complexity, memory and running time of these techniques in the table 2 shown below:

Table 2: Measures of Dimension reduction techniques

S. No	Technique	Computational Complexity	Running time(sec.)	Memory
1	PCA	$O(D^3)$	2.41	$O(D^2)$
2	Sammon Mapping	$O(n^3)$	30.51	$O(n^3)$
3	Isomap	$O(n^3)$	22.40	$O(n^2)$
4	Kernal PCA	$O(n^3)$	27.45	$O(n^2)$
5	Diffusion Map	$O(n^3)$	25.77	$O(n^2)$
6	Auto Encoders	O(inw)	675.29	O(w)
7	LLE	$O(pn^2)$	12.62	$O(pn^2)$
8	Laplacian Eigenmap	$O(pn^2)$	10.47	$O(pn^2)$
9	Hessian LLE	$O(pn^2)$	16.33	$O(pn^2)$
10	LLC Mani. Charting	$O(imd^3)$	10.09	$O(nm^d)$

8. Conclusion and Future work

In this paper we proposed a more secure, reliable and authenticate quantum cryptography based technique followed by convex and non-convex dimension reduction technique. Generally it is very difficult to implement to quantum cryptography due to high overhead of radiation problem of photon as well as physical implementation of Photon Light Pulse (PLP). So in future we can implement proposed model with QDK and use more efficient non-linear dimension reduction technique whose complexity and running time is lesser than the present techniques. This technique can also introduce a more secure and authenticate communication channel for community and blog creation.

9. References

[1] W.E. Arnoldi. The principle of minimized iteration in the solution of the matrix eigenvalue problem. Quarterly of Applied Mathematics, 9:17–25, 1951.

[2] Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning a Mahalanobis metric from equivalence constraints. Journal of Machine Learning Research, 6(1):937–965, 2006.

[3] S.P. Boyd and L. Vandenberghe. Convex optimization. Cambridge University Press, New York, NY, USA, 2004.

[4] E.W. Dijkstra. A note on two problems in connexion with graphs. Numerische Mathematik, 1:269–271, 1959.
[5] P. Demartines and J. H´erault. Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. IEEE Transactions on Neural Networks, 8(1):148–154, 1997.
[6] D.L. Donoho and C. Grimes. Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data. Proceedings of the National Academy of Sciences, 102(21):7426–7431, 2005.
[7] R. Duraiswami and V.C. Raykar. The manifolds of spatial hearing. In Proceedings of International Conference on Acoustics, Speech and Signal Processing, volume 3, pages 285–288, 2005.
[8] Faloutsos and K.-I. Lin. FastMap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, pages 163–174, New York, NY, USA, 1995. ACM Press.
[9] K. Fukunaga. Introduction to Statistical Pattern Recognition. Academic Press Professional, Inc., San Diego, CA, USA, 1990.
[10] G.E. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. Neural Computation, 18(7):1527–1554, 2006.
[11] H. Hoffmann. Kernel PCA for novelty detection. Pattern Recognition, 40(3):863–874, 2007.
[12] J.A. Lee and M. Verleysen. Nonlinear dimensionality reduction of data manifolds with essential loops. Neurocomputing, 67:29–53, 2005.
[13] J.A. Lee and M. Verleysen. Nonlinear dimensionality reduction. Springer, New York, NY, USA, 2007.
[14] A.M. Posadas, F. Vidal, F. de Miguel, G. Alguacil, J. Pena, J.M. Ibanez, and J. Morales. Spatial-temporal analysis of a seismic series using the principal components method. Journal of Geophysical Research, 98(B2):1923–1932, 1993.
[15] N. Gisin et al, “Quantum cryptography,” Rev. Mod. Phys., Vol. 74, No. 1, January 2002.
[16] Elliott, “Building the quantum network,” New J.Phys. 4 (July 2002) 46.
[17] G. Brassard and L. Salvail, “Secret key reconciliation by public discussion,” Lect. Notes in Computer Science 765, 410. (1994).

Mr. Lokesh Jain. is a M.Tech.(IS) final year student in Netaji Subhas Institute of Technology, Delhi, India. He has completed his graduation in 2005 from UPTU, India with honors. He has been associated with many academic institutes since last 4 years as faculty. He has been published more than 10 national and international papers on database and social networking. His area of interest is Database, Machine Learning, Discrete Mathematics and fuzzy Logic. Presently he is working on the social network security and privacy project.

Dr. Satbir Jain. has been associated with many international societies such as IEEE, CSI and IETE. He has been published more than 50 international and national papers. He is also a member of editorial committees of many technical societies. His area of interest is Database, Datamining, Data Modeling, Object Orientation, Software Engineering and Image Processing. Presently Dr. Jain is a professor in Department of Computer science in Netaji Subhas Institute of Technology, Delhi, India.