TABLE OF CONTENTS

1. INTRODUCTION

1.1. TITLE

Analysing the evolution of communication patterns in email data through an extended dynamic network analysis toolkit

1.2. RESEARCH QUESTIONS

How can similarity measures be used to analyse dynamic email networks? How can similarity measures in alternative spaces support the analysis of dynamic networks?

1.3. PURPOSE

The origins of graphs theory can be traced back to Leonhard Euler and his approach to solving the Konigsberg Bridge Problem. This city was located on the Pregel River in Prussia. The river divided this city into 4 distinct areas which included an island all of which were connected by a total of 7 bridges. Euler's representation of this problem of the individual areas as nodes and the bridges as edges is considered one of the first applications of graph theory. [1]

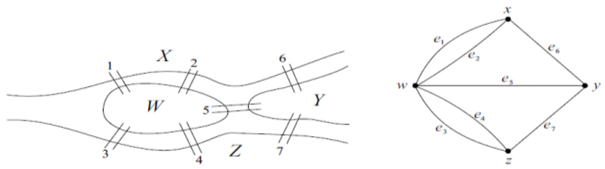


Figure 1: Euler's graphical representation of the Konigsberg Bridge Problem. [1]

Using Euler's insights into this problem in modern time's graphs have become a powerful tool with which to model and analyse communication and information networks. More specifically communities and their evolution in time [2]. The aim of this project is to explore the use of similarity measures to characterise such change in these networks. The viability of such analysis has been well demonstrated in the literature more specifically by [3], who modelled the email data released during the Enron investigation as a network and have made it available freely. Using this dataset [4] propose novel methods for community detection in real world information networks.

We intend to add to the existing body of knowledge regarding graph similarity measures and contribute a comparative and evaluative analysis of such measures on the emerging field of dynamic network analysis.

1.4. OBJECTIVES

The objectives that will help us answer our research question are as follows:

- What are some of the graph similarity measures that have been proposed in the literature?
- How have these been used in a practical context?
- How can such measures be applied to dynamic networks?
- How are such measures evaluated?

The aims of this project can be summarised as follows:

- To provide a fairly comprehensive overview of graph theory as is relevant to the understanding of the derivation of similarity measures.
- Conducting an in depth literature review on the graph similarity measures proposed and that have been demonstrated to be useful in a practical context.
- Compare the utility and performance of these measures on appropriate data
- Explore the viability of developing a novel similarity measure based on Fourier and Hilbert Analysis of networks

1.5. MOTIVATIONS & AIMS

The purpose of this research is twofold: firstly, we demonstrate the practical need for such analysis and secondly to develop a reference document for end user interested in applying such analysis to domains other than our stated domain of email networks.

The issue of graph similarity is critical in the analysis of dynamic networks. This is because when considering such networks an important concern is to understand the evolution of the network through time. Utilising similarity measures represents a very convenient method of achieving this as we can compare the same measure at different points in time to get a sense of network evolution. This is necessary because within our chosen context of email data, these methods can give insight into how communications patterns have evolved over time. Also we can gain insight into communities within such data and their birth, death and evolution. It can also serve as a means of anomaly detection. The set of analysis that such measures can enable are limited by one's creativity. The purpose is to show that there are very practical and real need to address the extension of similarity analysis to dynamic network analysis. We present many methods proposed in the literature and will also build on recent work on the application of Signal Processing methods on graphs to assess the potential of novel similarity methods.

The aim is to provide a comprehensive summary of graph similarity measures that can be applied to dynamic networks. We will explore their theoretical basis and methods of evaluation. In addition to the above it is hoped that the material developed as part of this project will provide a useful base reference to anyone interested in applying graph analytics to solve problems.

Also, we will explore the extent to which the measures that we encounter are implemented in common analysis packages such as IPython [5], NetworkX [6], Gephi [7] and Cytoscape [8] among others. The key point to make regarding tools selected here are that they are open source so they are freely available and the source code is viewable. This will allow if the need arises for extensions to be made to the code and shared freely.

2. CRITICAL CONTEXT

2.1. LIMITATIONS OF TRADITIONAL SNA

In Social Network Analysis (SNA), traditionally bounded networks are considered with maybe 2 or 3 connection or link types such as friendship or advice between a node types such as people sometimes another node type such as events are also considered together. [9]

If we consider more critically the interactions possible within our problem context of email networks we can have email networks within an organisation which are bounded and also with other organisations, clients and stakeholders and then the network does become unbounded. These networks can then be thought of as a higher order networks and as [9] notes many tools developed for simpler networks do not scale well to increased network size and complexity and in some cases experience degradation through increased susceptibility to Type 1 and Type 2 errors.
The dynamics in these networks can arise from different processes depending on the context of the problem. Natural evolutionary processes would be learning, births, deaths and ageing Others could be as a result of intervention measures such removal or addition of nodes i.e. removing those who lead the system, communities forming or disintegrating. The data associated with such systems are also often incomplete and contain errors which make the process of analysis and evaluation of these systems. [10]

Analysis approaches that go beyond traditional SNA and link analysis are therefore necessary. Within the context of such dynamic networks analysis can be performed to identify of key individuals, locating hidden groups and estimate performance. The data analysis process on such networks then involve: [9], [10]

- Relationship identification among nodes
- Network structure characterisation
- Locating the elite within the network
- Identifying points of vulnerability
- Comparing networks

The approaches that enables effective analysis of such dynamic networks and help quantify their evolution over time is the motivation for this research.

2.2. GRAPH TERMINILOGY

A graph, G can be described as a triple which consists of a set of edges E (G), a set of vertices or nodes V

(G) and a relationship that connects the vertices to these edges. Finite graphs are those that have V and E as a finite set. Simple graphs are those that have no loops or multiple edges. A path is simple graph in which the vertices can be ordered where two vertices can be adjacent only if they are consecutively ordered. A cycle is defined as a simple graph where the vertices can be cyclically ordered such that two vertices are adjacent only if they are consecutive in cyclical ordering. A subgraph can be thought of cycles and paths within a larger graph, where the edge relations between the subgraph and the large graph are the same. [1]

2.3. DYNAMIC NETWORK ANALYSIS AS AN EXTENSION TO SNA

Dynamic network analysis (DNA) aims to extend the methods, tools and techniques used in traditional Social Network Analysis (SNA) to the analysis of networks which are able to handle big dynamic multi-mode, multi-link networks with varying levels of uncertainty. Dynamic networks also allow for probabilistic connection between nodes. [9]

In [9] DNA was explored within the context of terrorism networks. Here an additional layer of complexity is added by the fact that an act of measurement changes its properties and this change propagates through the network and its state changes. Another key point is that the nodes in this network have the ability to learn. So the nodes themselves can be thought of being probabilistic compared to the more static nature of SNA nodes.

In a DNA representation systems can be represented as relational data. This relational data structure can lends flexibility in defining multiple node types defined as multi-modal, have various types of connections among such nodes called multi-plex. The underlying attributes of both node, edges and the data change over time hence the dynamic part. [10]

In [9] the key advances that allow for the analysis of such dynamic networks are identified as:

- ⊕ The meta matrix
- ⊕ Probabilistic edges between nodes
- ⊕ Combining social networks with cognitive science and multi-agent systems

The Meta Matrix

The Meta matrix is a method used in operations research and organisational management that seeks to represent the entity and class relationships as a collection of networks. In the DNA context this translates as a multi-mode, multi-plex approach to representing systems. Therefore the Meta matrix can contain a social network, a membership network and knowledge network and allow us to explore and analyse the connections between them. [9]– [12]

Probabilistic Ties

The ties or connections in the Meta matrix are probabilistic with various factors affecting their probability. This allows for inclusion of the observers uncertainty and the likelihood that the tie is present at the time of observation. These probabilities themselves and their temporal evolution maybe estimated by the Bayesian methods, cognitive inferencing and models of social and cognitive change. [9]

Multi Agent Network Models

As previously discussed the SNA treatment of nodes as static agents unable to learn is insufficient when dynamic networks are concerned. In DNA the nodes are able to take actions, learn from experience and alter their networks as a result. Some social and cognitive processes that influence the agent's interactions are relative similarity, relative expertise and co-workers. The dynamic behaviour of the network emerges from these interactions and experience a shared evolution. [9]

We briefly discuss some of the more common measures associated with networks which relate to their global and local properties. These will be important when we discuss similarity because one of the ways to assess similarity is to consider snapshots of a network attribute at different time intervals.

2.4. NETWORK MEASURES: LOCAL

Centrality measures are a fundamental statistic in network analysis. In [13] two paradigms of centrality definitions are suggested. One is the means based definition of centrality or the graph theoretic and the other is the ends-based definition which is a dynamic model based view that focuses on the outcome for the nodes in a network where there is flows across the nodes [14]. However, both approaches agree that this measure is a node level property.

In both formulations of centrality measures noted above we characterise centrality measures as follows: [13]

- ⊕ Volume based measures – degree like centrality
- ⊕ Length based measures – closeness like centrality
- ⊕ Medial measures – betweenness like centrality

Volume and length based measures are what are called radial measures because they analyse walks that emanate from or terminate with a given node. Medial measures on the other hand are based on position so how many times does one encounter a node while trying to reach other nodes in the network.

Degree Centrality
Degree Centrality is a special case of the k-path centrality that counts the all the paths of a length, k that originate from a given node. Degree Centrality is then defined as the number of edges incident on a node. Which translates to summing all the rows of the adjacency matrix of a network. [13], [14]

Closeness Centrality
Closeness Centrality is the graph theoretic distance or the geodetic distance from a given to all the other nodes in a network. [13], [14]

Betweenness Centrality
Betweenness centrality counts the number of times that a certain node, x needs to pass another node, y to get to another node, z through the shortest path between them. [13], [14]

Eigenvector Centrality
The Eigenvector Centrality is defined as the principal eigenvector of the adjacency matrix of a network. It captures the intuition that nodes that have high eigenvector centrality scores are likely to be close to other nodes which themselves have high values for this measure. [13], [14]

We have mentioned some of the most popular centrality measures but there are numerous other variations mentioned in the literature and are beyond the scope of this work the interested reader is referred to the following starter references. [15]– [18]

2.5. NETWORK MEASURES: GLOBAL

Network Size
This can be defined by counting the number of nodes in a network. [19]

Density
This is defined as the total number of edges divided by the total number of possible edges. [19]

Diameter
The diameter of a network is the maximum geodesic distance between two nodes.

Modularity
The modularity function finds partitions within the graph where a large proportion of the edges fall entirely within that partition and biases against those partitions that have too few or unequal sized parts. [20]

3. APPROACHES

The purpose of detailing the some of the key properties and attributes are they can be used as to derive similarity measures on graphs. These measures will form the basis of our exploration of existing and new similarity measures for DNA.

The problem of graph similarity or graph matching then becomes one of finding the equivalence of two graphs with potentially different number of nodes and edges and returning a measure within [0, 1] that captures their similarity or dissimilarity. [18], [21]– [27]

The key idea of graph matching in the context of dynamic networks can be summarised as finding a subgraph or an attribute that we can compare between two time instances. For example, if we consider the Degree Centrality of a network at time step 0 and then again calculate this measure at time step 1 we can apply a similarity measure on this attribute to quantify the change within the network. This can be done by means of a distance metric such as cosine similarity and others are possible.
Therefore the analysis methodology can be generalised as follows:

- ⊕ For network X at time step, **t** extract some attribute
- ⊕ At time step, **t +1** extract the same attribute
- ⊕ Perform similarity analysis on attribute
- ⊕ Track the change in the similarity metric over time through some control process.

The evaluation of the change in metrics over time will be done through a statistical control process as suggested in [28]. This is a concept that comes from quality engineering and it essentially involves calculating a statistic from a sequence of measurements of a random process and then comparing it so some control limit. This process translates to:

- Calculating a cumulative sum control chart which is very good for detecting small changes in mean over time
- Calculating a z-score for each time step (Eq. 1)
- Construction of two charts to detect increase and decrease in the metric as shown.

$$Z_i = (x_i - \mu_0)/\sigma$$
$$C_t^+ = \max\{0,\ Z_t - k + C_{t-1}^+\}$$
$$C_t^- = \max\{0,\ -Z_t - k + C_{t-1}^-\}$$

Equation 1: Control chart to detect increase and decrease in metric over time

In the following section will present a brief treatment of the subject of similarity.

3.1. SIMILARITY METHODS

The problem of finding similarity between nodes which are similar in a network can be thought of as a problem of finding a set of nodes which are similar to a given node according to some attributes which are represented as connections [25].

Similarity in a networks is classified as being of structural, content or keyword based The Structural similarity or link based similarity considers the similarity of links between the nodes in the graph e.g. Cosine, Jaccard, Hub Promoted and Hub Depressed Index etc. Content similarity considers the attributes of the node in the graph. For example on a social network this could be birth dates or hobbies of individuals. Keyword similarity aims to find similarity based on nodes representing word collections. Global Structural Similarity can be classified as being:

- local vs. global
- parameter-free vs. parameter-dependent
- node-dependent vs. path- dependent

The global structural measures aim to measure node similarity compared to the whole network. We will call them intra network similarity measures. [25]

Inter network similarity measures are described in [22], [24], [26], [27], [29]– [32]. These measures are classified by [21] into three categories 1) Distance Based 2) Feature Based and 3) Probabilistic.

Distance based approach
The distance based approach is perhaps the earliest of the methods encountered which is based on edit distance [32]. Essentially this boils down to finding a sequence of operations such as deletion, insertion, or substitution minimising some cost function that will turn one graph into another. These involve detection and comparison of the graph isomorphism, subgraph isomorphism and maximum common subgraph detection utilising the edit distance. Although these methods are guaranteed to converge to an optimal solution their exponential complexity makes them unsuitable for large graphs.

Feature Based approach
We have already hinted at the feature based approach above in the Al-Qaeda example above. But more formally this involves calculation of a network attribute such as degree, closeness, betweenness, and/or eigenvector centrality for the graphs and then applying a similarity measure on them that will characterise their similarity or dissimilarity. This has the benefit of being scalable to very large networks as the aggregated statistics are much smaller than the network themselves.

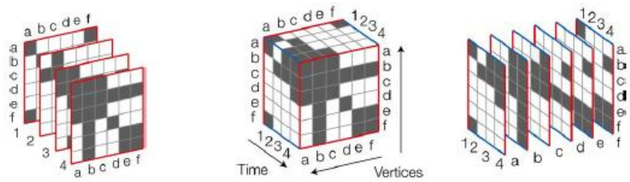
A taxonomy of the methods that have been proposed to solve these problems are shown in Fig 3. We include this for illustrative purposes and their discussion will form part of the extended review performed for the final report.

Probabilistic Approach
The methods that fall under this approach in the literature are vast. Some approaches under the probabilistic framework for graph matching are discussed here [29], [33]– [35]. But simply stated these methods define a probability distribution over mappings or graph embedding's [31], [34]. Graph embeddings are graphs whose nodes correspond to distinct points on a plane and the edges represents relationships connecting these points. The matching algorithm is strongly dependent upon the geometric information attached to the graphs [31], [34].

Graph matching allows for recovering point correspondences. In [29] the authors show that assuming that the assignment matrix that represents these correspondences are statistically independent the high order matching problem can be represented by a Kronecker product matrix. Also they show that

that a high order tensor affinity tensor can be marginalised into a one dimensional vector of probabilities. This probability vector is then updated by projection to a vector assignment space and then minimizing a distance measure (Bregman measure) [33]. Spectral Methods involve looking at the spectra of a graph which are defined by the eigenvalues of the adjacency matrix. In [36] the high order matching is expressed a tensor Eigen decomposition and applied to point matching using some similarity measure.

Visual Representation
More recently, the authors in [37] have proposed visualising dynamic networks and characterising change by visualising the adjacency matrix of these networks as a matrix cube. Representing the adjacency matrix as a stack of cubes rather than node link diagrams is found to be a much more useful paradigm for analysis of dynamic networks especially when these networks are dense.



(a) Adjacency Matrices (b) Matrix Cube (c) Vertex Slices

Figure 2: The matrix cube view of dynamic networks that allow for visual characterisation of network evolution over time. [37]

We will explore some representative methods from these classes of methods and assess their utility for DNA.

3.1. ORIGINAL IDEAS

Given the author's background as a Geophysicist in performing seismic data analysis, there was a strong interest in exploring potential ideas that could be imported from seismic analysis to the analysis of dynamic networks. The reasons for this is because seismic data is essentially a time series on which signal processing approaches are utilised. Since there is a time component to these networks the attributes of these networks could be viewed as a time series and some of these methods can then be tested for effectiveness.

Firstly, for visualisation of dynamic network attributes we suggest something similar to how well log analysis is performed in reservoir characterisation of hydrocarbon reservoirs shown in Fig 3.

From Fig 6, we propose to develop a dynamic network attribute panel that allows analysis of selected network attributes over time such as Degree, Closeness, and Density etc. [38]

The Coherence attribute is multi trace similarity of the seismic waveform that is used for event detection and is very useful when the amplitude of the signal is low compared to the noise. For DNA we can explore the viability of developing a Coherence for network similarity measures or network attributes that also allows for event detection shown in Fig 7. More details regarding this method can be found here [39].

Some recent work has focused on the application of signal processing methods to graphs. In [40] a generalisation of the Fourier Transform and its associated methods such as convolutions to graphs. In [41], [42] transforms and tomograms for graph application are discussed. An interesting application is that of the Radon transform commonly used for noise removal in seismic data analysis, which the authors in [42] generalise to arbitrary pairs of non-commuting operators. These are positive bilinear transforms with a probabilistic interpretation and provide a full characterization of the signals while being robust to noise. These developments although very interesting do not have mainstream implementations in software packages and must be conducted in a more ad-hoc manner. This is a potential limitation of their evaluation in our study.

However, we present some preliminary analysis on a synthetic geographical threshold graph model which places n nodes uniformly at random in a rectangular domain with 500 nodes. We extract the Degree Centrality and then apply the Fourier and Hilbert Transform to it to extract its periodic features and complex trace information.

From Fig 3, 4 we note that these transforms serve as a useful visualisation tool of network attributes. The Fourier Transform picks out the maximum degree centrality while the Double Hilbert Transform shows some outliers.

A correlation between the Fourier or Hilbert transforms of a measure would give a similarity value shown in Fig 8. The autocorrelation of the measures with themselves would give an indication of the self-similarity of the network.

From the autocorrelation functions we see that these have the potential to highlight anomalies. We will explore these measures over time to assess their value in DNA.

In summary, our analysis approach will focus on network measures calculated through some of the methods we identified previously. We will consider the temporal component as an important aspect of network characterisation.

The key consideration of the analysis will be to use already implemented method and not be distracted by extensive development work as this is not the purpose of the research. Hence, we identify promising areas to focus on but if time constraints do not permit or the software packages are found to be lacking we will include them as future areas of research.

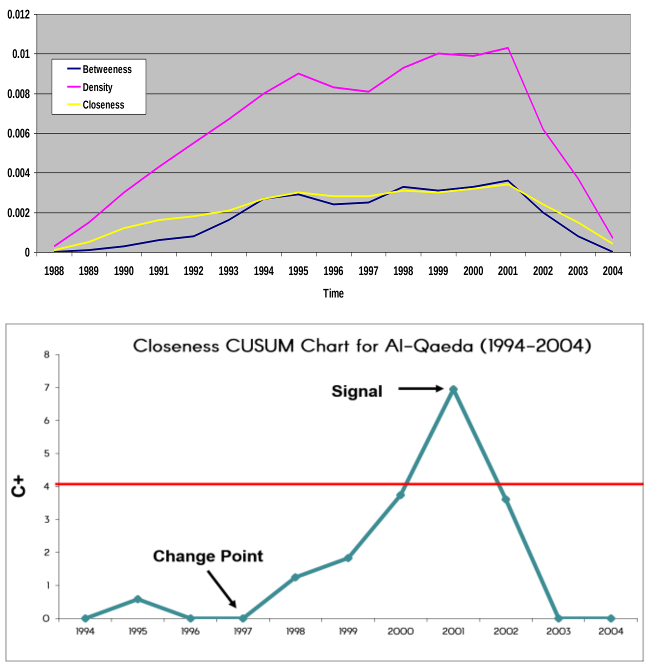


Figure 5: From [28] (top) shows the application of this method on the Al Qaeda network for the Betweenness, Density and Closeness measures. Bottom modified from [28] shows the change in Closeness Centrality for the network through the control chart. The control limit is shown in red

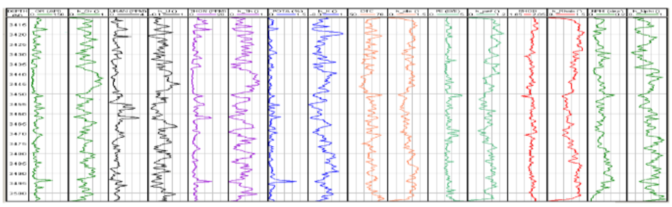


Figure 6: Well Log panel showing the different measurements typically recorded while drilling. These attributes are then used for characterisation.

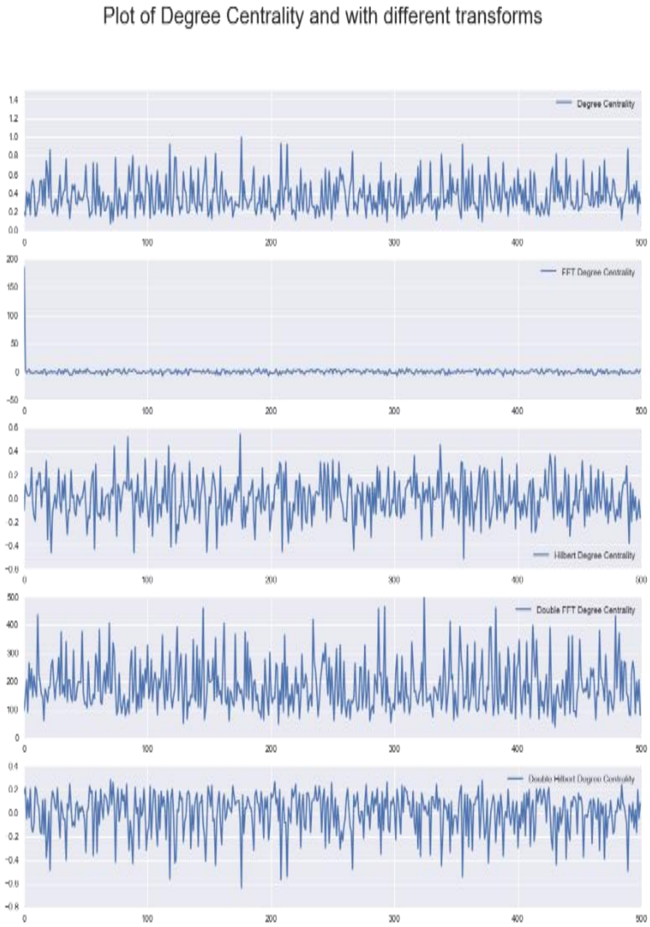


Figure 3: Plot of (1st) Degree Centrality (DC), (2nd) Fourier Transform of DC, (3rd) Hilbert Transform of DC, (4th) Double Fourier Transform of DC, (5th) Double Hilbert Transform of DC

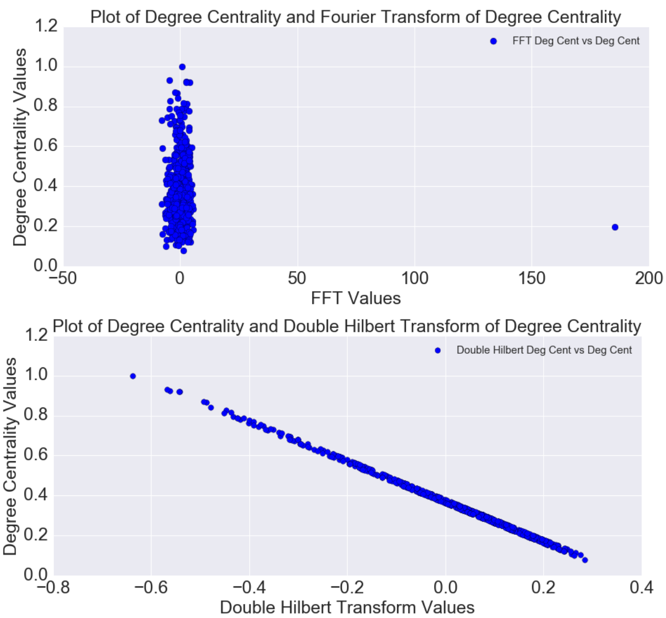


Figure 4: (top) Scatter plot of Fourier Transform of DC and DC (bottom) Double Hilbert Transform of DC and DC.

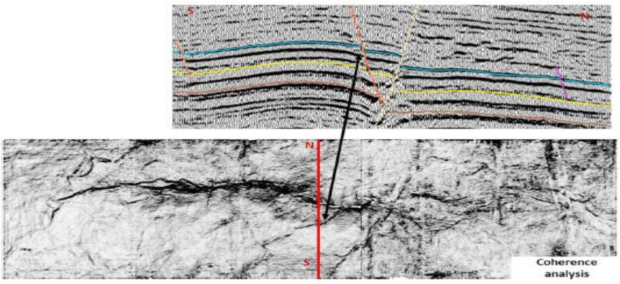


Figure 7: Example of a coherence attribute cube for a seismic volume. Coherence is essentially a multi trace similarity measure used for event detection. [39]

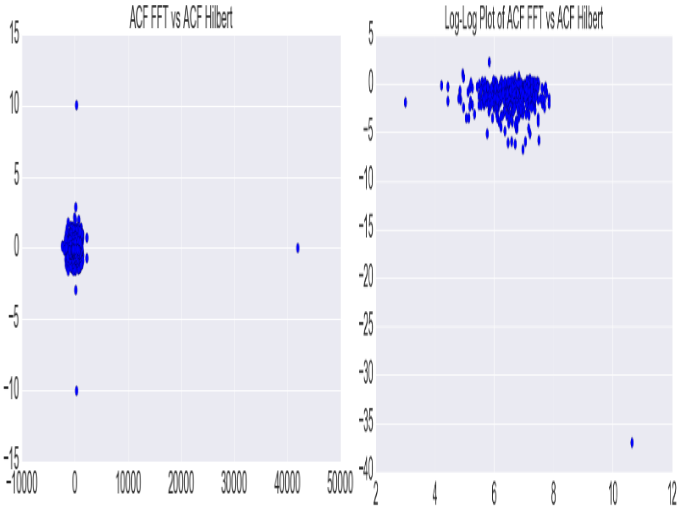


Figure 8: Plot of autocorrelation functions of the Fourier and Hilbert Transforms of Degree Centrality. (Left) Log-Log Plot of ACF's.

3.2. EVALUATION STRATEGIES

We have presented an array of methods and techniques proposed in the lecture and our own novel ideas based on the literature. Evaluation of such measures and their interpretation will form a key part of the research. The interpretation will be especially important when we consider our suggestions of measures of similarity in alternative spaces such as the Fourier and Hilbert spaces. These Fourier Transform extracts periodic features from a time series by extracting its frequency while the Hilbert transform extends the signal to the complex plane. What these mean in the context of our measures will be explored.

Therefore the evaluation strategy will be as follows:

- For the dataset used search the literature to see what other authors have found especially with regards to interesting features, signals and/or structures
- Then evaluate whether our measures are able to find these features and how difficult or easy they are in comparison with the original measures
- If such signals are not found in the literature, we can look for them in the data through other analysis and use our measures to confirm alternatively we can insert a test signal in the data and see how well the measures are able to find them.

This is deemed the most practical way to evaluate the research given the short time horizon. Other methods such as involving groups or surveys to assess participant's interactions with these analytical tools would not be practical for this study.

3.1. DATA SOURCES

This research will utilise open and publicly available data so the results are reproducible. To this end we identify some sources of network data available to us:

- Enron email data : http://www.cs.cmu.edu/~enron/
- Stanford Large Network Collection: http://snap.stanford.edu/data/#email
- UCI Network Data Repository : https://networkdata.ics.uci.edu/resources.php
- Open Dynamic Network Data: http://www.aviz.fr/~bbach/opendynamicnetworks

4. WORK PLAN

The research will be conducted between the periods of June – September 2016. A detailed work plan is highlighted in the Gantt chart attached with this report.

5. ETHICS

Since this is predominantly a research and analysis task on open data there are ethical concerns. Completed ethics questionnaire is also appended with this report.

6. RISK REGISTER

The standard types of technical and non-technical risks apply to this project as any other. A completed risk register is appended with this report.

7. REFERENCES

[1] A. Dickson, "Introduction to Graph Theory The Konigsberg Bridge Problem," *October*, no. October, pp. 1–15, 2006.

[2] A. Chapanond, M. S. Krishnamoorthy, and B. Yener, "Graph theoretic and spectral analysis of Enron email data," *Comput. Math. Organ. Theory*, vol. 11, no. 3, pp. 265–281, 2005.

[3] B. Klimt and Y. Yang, "Introducing the Enron Corpus," *Mach. Learn.*, vol. stitutep1, p. wwceasccaers2004168, 2004.

[4] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters," *Internet Math.*, vol. 6, no. 1, pp. 29–123, 2011.

[5] F. Pérez and B. E. Granger, "IPython: a System for Interactive Scientific Computing," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 21–29, May 2007.

[6] H. Aric, D. Schult, and S. Pieter, "NetworkX." .

[7] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks," *Third International AAAI Conference on Weblogs and Social Media*, 2009. [Online]. Available: http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154\npapers2://publication/uuid/CCEBC82E-0D18-4FFC-91EC-6E4A7F1A1972. [Accessed: 25-Mar-2016].

[8] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: A software Environment for integrated models of biomolecular interaction networks," *Genome Res.*, vol. 13, no. 11, pp. 2498–2504, Nov. 2003.

[9] K. M. Carley, "Dynamic Network Analysis," no. September, 2001.

[10] K. M. Carley, J. Diesner, J. Reminga, and M. Tsvetovat, "Toward an interoperable dynamic network analysis toolkit," *Decis. Support Syst.*, vol. 43, no. 4, pp. 1324–1347, 2007.

[11] K. M. Carley, "Ora: Organization risk analyzer," *Science (80-.).*, no. January, p. 49, 2004.

[12] T. A. B. Snijders, "Introduction To Dynamic Social Network Analysis," *BRiMS 2015*, no. November, 2015.

[13] S. P. Borgatti and M. G. Everett, "A Graph-theoretic perspective on centrality," *Soc. Networks*, vol. 28, no. 4, pp. 466–484, 2006.

[14] S. P. Borgatti, "Centrality and network flow," *Soc. Networks*, vol. 27, no. 1, pp. 55–71, 2005.

[15] G. Costantini, S. Epskamp, D. Borsboom, M. Perugini, R. Mõttus, L. J. Waldorp, and A. O. J. Cramer, "State of the aRt personality research: A tutorial on network analysis of personality data in R," *J. Res. Pers.*, vol. 54, pp. 13–29, 2015.

[16] K. R. Harrison, M. Ventresca, and B. M. Ombuki-Berman, "A meta-analysis of centrality measures for comparing and generating complex network models," *J. Comput. Sci.*, 2015.

[17] R. Kaur and S. Singh, "A survey of data mining and social network analysis based anomaly detection techniques," *Egypt. Informatics J.*, 2015.

[18] R. a. Hanneman and M. Riddle, "Measures of similarity and structural equivalence," *Introduction to social network methods*, 2011. [Online]. Available: http://www.faculty.ucr.edu/~hanneman/nettext/C13_Structural_Equivalence.html. [Accessed: 26-Mar-2016].

[19] R. a Hanneman and M. Riddle, "Introduction to Social Network Methods," *Riverside, CA Univ. California, Riverside. On-line Textb.*, vol. 46, no. 7, pp. 5128–30, 2005.

[20] C. McDiarmid and F. Skerman, "Modularity in random regular graphs and lattices," *Electron. Notes Discret. Math.*, vol. 43, pp. 431–437, 2013.

[21] F. Ashby and D. Ennis, "Similarity measures," *Scholarpedia*, vol. 2, no. 12, p. 4116, 2007.

[22] M. Dehmer, F. Emmert-Streib, and J. Kilian, "A similarity measure for graphs with low computational complexity," *Appl. Math. Comput.*, vol. 182, no. 1, pp. 447–459, Nov. 2006.

[23] M. Komosinski and M. Kubiak, "Quantitative measure of structural and geometric similarity of 3D morphologies," *Complexity*, vol. 16, no. 6, pp. 40–52, 2011.

[24] D. Koutra, a Parikh, a Ramdas, and J. Xiang, "Algorithms for Graph Similarity and Subgraph Matching," 2011.

[25] A. Rawashdeh and A. L. Ralescu, "Similarity Measure for Social Networks – A Brief Survey," 2012.

[26] S. Soundarajan, T. Eliassi-Rad, and B. Gallagher, "A guide to selecting a network similarity method," *Sdm*, no. 1, 2014.

[27] L. A. Zager and G. C. Verghese, "Graph similarity scoring and matching," *Appl. Math. Lett.*, vol. 21, no. 1, pp. 86–94, 2008.

[28] "Dynamic Social Network Analysis of Al-Qaeda." .

[29] R. Zass and A. Shashua, "Probabilistic graph and hypergraph matching," *26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, pp. 0–7, 2008.

[30] C. Wang, L. Wang, and L. Liu, "Improving graph matching via density maximization," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 3424–3431, 2013.

[31] D. Conet, P. Foggia, C. Sansone, and M. Vento, "Thirty Years of Graph Matching in Pattern Recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 18, no. 03, pp. 265–298, 2004.

[32] H. Bunke, "Recent developments in graph matching," *Pattern Recognition, 2000. Proceedings. 15th ...*, vol. 2, pp. 117–124, 2000.

[33] A. Egozi, Y. Keller, and H. Guterman, "A Probabilistic Approach to Spectral Graph Matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 18–27, 2013.

[34] T. S. Caetano, J. J. McAuley, L. Cheng, Q. V. Le, and A. J. Smola, "Learning graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1048–1058, 2009.

[35] H. Yaghi and H. Krim, "Probabilistic graph matching by canonical decomposition," *Image Process. 2008. ICIP 2008. 15th IEEE Int. Conf.*, pp. 2368–2371, 2008.

[36] O. Duchenne, F. Bach, K. In-so, J. Ponce, O. Duchenne, F. Bach, K. In-so, J. Ponce, A. T. Algorithm, O. Duchenne, F. Bach, I. Kweon, and J. Ponce, "A Tensor-Based Algorithm for High-Order Graph Matching," 2014.

[37] B. Bach, E. Pietriga, and J.-D. Fekete, "Visualizing dynamic networks with matrix cubes," *32nd Annu. ACM Conf. Hum. Factors Comput. Syst. CHI 2014*, pp. 877–886, 2014.

[38] L. Aliouane, S. Ouadfeul, and A. Boudella, "Well-Logs Data Processing Using the Fractal Analysis and Neural Network," 2012.

[39] C. Outline, "Seismic Coherence Technique," pp. 63–74, 2014.

[40] M. Begú, "Fourier Analysis on Graphs."

[41] M. A. Man'ko, R. V. Mendes, and V. I. Man'ko, "Tomograms and other transforms: a unified view," *J. Phys. A. Math. Gen.*, vol. 34, no. 40, pp. 8321–8332, 2001.

[42] A. Sandryhaila and J. M. F. Moura, "Discrete Signal Processing on Graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.