



**CITY UNIVERSITY
LONDON**

**MSc Data Science
Project Report
2015/16**

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

Arshad Ahmed

Supervised by: Dr Cagatay Turkey

Date of Submission:

Declaration

By submitting this work, I declare that this work is entirely my own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the assessment instructions and any other relevant programme and module documentation. In submitting this work, I acknowledge that I have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. I also acknowledge that this work will be subject to a variety of checks for academic misconduct.

Signed: *Arshad Ahmed*

Abstract

Table of Contents

1.	Introduction.....	5
1.1.	Background.....	5
1.2.	Social Network Analysis (SNA).....	6
1.3.	Dynamic Network Analysis (DNA).....	6
1.4.	Super networks Analysis (SA).....	6
1.5.	Motivations & Aims.....	7
1.6.	Objectives.....	7
2.	Critical Context.....	9
2.1.	The Limitations of traditional SNA.....	9
2.2.	Dynamic Network Analysis as an Extension to SNA.....	9
2.2.1.	The Meta Matrix.....	10
2.2.2.	Probabilistic Ties.....	10
2.2.3.	Multi Agent Network Models.....	10
2.3.	Network Measures: Local.....	10
2.3.1.	Degree Centrality.....	11
2.3.2.	Closeness Centrality.....	11
2.3.3.	Betweenness Centrality.....	11
2.3.4.	Eigenvector Centrality.....	11
2.4.	Network Measures:: Global.....	11
2.4.1.	Network Size.....	11
2.4.2.	Density.....	11
2.4.3.	Diameter.....	12
2.4.4.	Modularity.....	12
2.5.	Graph Similarity.....	12
2.6.	Similarity Methods.....	13
2.6.1.	Distance based approach.....	14
2.6.2.	Feature Based approach.....	14
2.6.3.	Probabilistic Approach.....	14
2.7.	Spectral and Tensor Methods.....	15
2.8.	Visual Representation.....	15
3.	Methods.....	16
4.	Results.....	17
5.	Discussion.....	18
6.	Evaluation, Reflections & Conclusions.....	19
7.	References.....	20

1. Introduction

1.1. Background

The origins of graphs theory can be traced back to Leonhard Euler and his approach to solving the Königsberg 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]

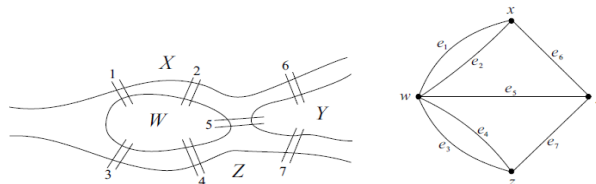


Figure 1: Euler's graphical representation of the Königsberg Bridge Problem.[1]

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]

Networks are all around us from the internet networks that enable us to connect to each other through a range of mediums, to the transport networks that facilitate movement of goods and people to communication networks that enable us to communicate easily across boundaries. They all share a key characteristic that they are complex systems that can be effectively modelled through the use of graph analysis or network modelling. [2]

In this report we use the terms graph and network interchangeably.

Graph analysis has myriads of applications such as in finance, where correlation matrices of returns for portfolios can be treated as an adjacency matrix and assets above a certain correlation can be assigned an edge between them. This allows for the data to be visualised as a network from which individual components which are not correlated and strongly connected subgraphs are easily distinguishable.[3]

In systems biology networks are used to integrate data from different studies and sources into a single framework using genes as anchors. These networks can also be directly be generated from time series or perturbation data and the changing quantities of an item of interest over time under different regimes can be studied using advanced methods such as Bayesian Networks.[4]

Tian *et al.*,2013 [2] approaches the task of Opinion Leader Identification from different paradigms of network analysis such as Social Network Analysis (SNA), Dynamic Network Analysis (DNA) and Super networks. Opinion leaders are entities who publish influential views, insights and arguments that guide public awareness on issues. The authors demonstrate through a case study that by building network models of interactions based on online

interactions such as blog posts, comments and their associated sentiments it is possible to identify such entities after a major event. This would be helpful for governments interested in identifying and silencing critics.

Chapanond et al., 2005 [5] applied SNA to identify structures within the Enron email dataset. They apply a combination of graph theoretic and spectral methods and discover that the email adjacency matrix has a rank 2 approximation and pre-processing of the raw data has a significant impact on the final conclusions. In addition, they find that the Enron graph some communities have a high number of links between them while other have lower numbers. They propose using relative interconnectivity to analyse community types. In conclusion they state that based on their chosen metrics of average distance ratio, compactness and clustering coefficient that the graph is well distributed, sparse and that the graph clusters are not dense.

A distinction needs to be made at this stage regarding the different types of network analysis that are possible which will inform the rest of this work. In this report we focus mainly on email network data but the methods can be generalised to other application areas.

For our context there are three relevant paradigms of graph analysis:[2]

- ⊕ Social Network Analysis (SNA)
- ⊕ DNA (DNA)
- ⊕ Super networks (SA)

1.2. Social Network Analysis (SNA)

SNA is concerned with the study of relationships between entities and its focus of research is of two types: whole network analysis and self-centred network analysis. Whole network analysis is concerned with understanding the structure of relationships between different roles in a group and is used to investigate network structure changes with the time and the contact pattern of network entities. Self-cantered network analysis is concerned with how the individual behaviour of network entities are influenced by the membership of the network. [2]

1.3. Dynamic Network Analysis (DNA)

DNA was proposed as an extension to the SNA and this is discussed in greater detail in Chapter 2. The strength of DNA is that it is able to handle large scale dynamic, multi-modal, multi-lateral network with various levels of uncertainty. The edges are probabilistic and the nodes behave like agents in a multi agent environment so this enables perturbations or changes in the network to propagate through the network and result in some global reconfiguration. The evolution of a network in The application of machine learning and multi agent modelling in the same environment is enabled by DNA's use of the meta matrix. [2]

1.4. Super networks Analysis (SA)

Super networks can be thought of as networks of networks that exist above and beyond existing networks. These have the characteristic of being multi-layered, multi-dimensional, multi-attributed and multi-levelled with additional features such as congestion and coordination. These have been applied predominantly in supply chain management, finance, traffic and ecology among others. These networks are analysed using either variational inequality and/or hypergraph theory. [2]

The suitability of network analysis of email networks the graph analysis method for application on email networks is well established [5]-[7]. In this study we propose to use SNA not for just static graphs but for networks evolving in time. The aim here is to not just detect community structure but describe its evolution in time. We want to be able to detect the

formation, merger and death of communities over time. In the context of email networks this analysis can reveal who is talking to whom in an organisation. It can also show trends in the communication between departments for example for a major product launch there should be a reasonable expectation that product, marketing and customer service teams should be in communication and that the frequency should increase. But if we find that the communication patterns hardly change or one department is completely isolated this indicates problems. This is because if the customer services are disconnected about discussions about new products they would not be able to do a good job of helping customers when problems arise. Therefore, this sort of analysis can be used to find and prevent such bottlenecks. Dynamic measures should be able to answer whether communication communities across departments are developing and merging over time or are they separating and becoming more isolated over time.

The purpose of this research is to firstly is to conduct a review of measures and methods that can be used to characterise networks. Then we will investigate how these measures can be extended to the analysis of such dynamic networks. Within the context of email networks, we will investigate which measures work best and their possible interpretations. So this report is intended to demonstrate the utility of these methods and inform of other methods that could potentially be applied by others looking to conduct such analysis to solve problems in their respective domains.

1.5. Motivations & Aims

As mentioned the motivation is to investigate measures for the analysis and characterisation of dynamic networks and from these evaluate which measures work best. These measures should aid in the tracking of evolution in a network such as addition and deletion of members, formation, merger and death of communities. This will be interpreted in the context of email network data. In an organisation setting it should allow managers to evaluate the communication structure of their organisations and identify where bottlenecks lie. These communication bottlenecks could cause miscommunications and lead to potential loss of revenue. Therefore, active evaluation of measure to encourage greater communication is required as well as methods to analyse changes in the network over time. This work seeks to improve on broad measures that are available for the characterisation of static networks and their extension to the analysis of dynamic networks.

In order to do so we explore the issue of graph similarity, which is critical in the analysis of dynamic networks. 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. 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.

1.6. 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 such as email data?
- ⊕ 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

2. Critical Context

2.1. The 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. [8]

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 [8] 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.[9]

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: [8], [9]

- ⊕ 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. 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. [8]

In [8] 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 system can be represented as relational data. This relational data structure can lend 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. [9]

In [8] 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

2.2.1. 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. [8]-[11]

2.2.2. 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. [8]

2.2.3. 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. [8]

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.

2.3. Network Measures: Local

Centrality measures are a fundamental statistic in network analysis. In [12] 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[13]. 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:[12]

- ⊕ 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.

2.3.1. 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. [12], [13]

2.3.2. Closeness Centrality

Closeness Centrality is the graph theoretic distance or the geodetic distance from a given to all the other nodes in a network. [12], [13]

2.3.3. 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. [12], [13]

2.3.4. 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. [12], [13]

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.[14]–[17]

2.4. Network Measures: Global

2.4.1. Network Size

This can be defined by counting the number of nodes in a network.[18]

2.4.2. Density

This is defined as the total number of edges divided by the total number of possible edges. [18]

2.4.3. Diameter

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

2.4.4. 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.[19]

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.

2.5. Graph Similarity

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. [17], [20]–[26]

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 [27]. 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

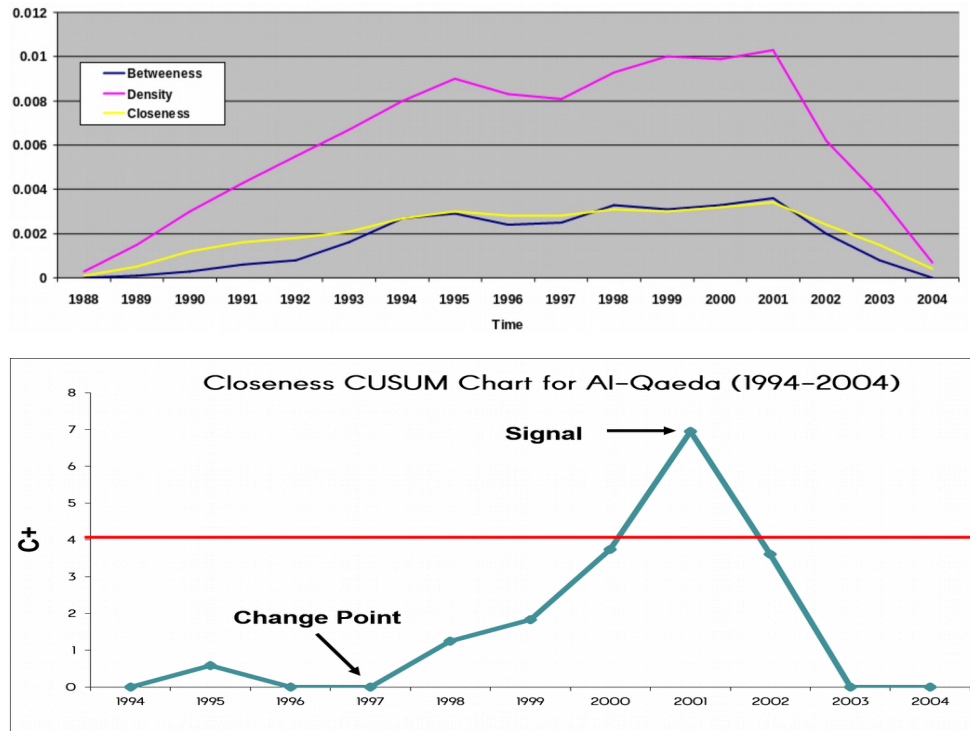


Figure 2: From [27] (top) shows the application of this method on the Al Qaeda network for the Betweenness, Density and Closeness measures. Bottom modified from [27] shows the change in Closeness Centrality for the network through the control chart. The control limit is shown in red

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

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[24].

2.6. Similarity Methods

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. [24]

Inter network similarity measures are described in [21], [23], [25], [26], [28]–[31]. These measures are classified by [20] into three categories:

- ⊕ Distance Based
- ⊕ Feature Based
- ⊕ Probabilistic

2.6.1. Distance based approach

The distance based approach is perhaps the earliest of the methods encountered which is based on edit distance [31]. 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.

2.6.2. 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.

2.6.3. 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 [28], [32]–[34]. But simply stated these methods define a probability distribution over mappings or graph embedding's [30], [33]. 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 [30], [33].

Graph matching allows for recovering point correspondences. In [28] 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) [32].

2.7. Spectral and Tensor Methods

Spectral graph theory is the study of the eigenvectors and eigenvalues of graph matrices [35]. The spectrum of a finite graph is the spectrum of the adjacency matrix, which is the eigenvectors and eigenvalues derived from the eigendecomposition of this matrix. For an undirected graph without loops, the Laplace Matrix is the matrix indexed by a vertex set of \mathbf{V} , with zero row sums if D is the degree matrix of a graph and A is the adjacency matrix then the Laplacian Matrix, L can be defined as $L = D - A$, where $Q = D + A$ is called the signless Laplacian Matrix of the graph. [35]-[37]

Spielman, 2007 [35] note that since the eigenvalues of a graph do not depend on the vertex ordering of the graph then they could be used to distinguish between pairs of non-isomorphic graphs. If the eigenvalues for the graphs are different then two graphs can be considered different. But he notes that this approach has problems such as the eigenvectors being only determined up to sign i.e. \mathbf{v} and $-\mathbf{v}$ can both be eigenvectors, so spectral embedding comparison would result in having to check 2^k possible ways of flipping their signs. The eigenvectors can provide coordinates for each vertex in a graph which is independent of the vertex labels but for graphs for which non-trivial eigenvalues have high multiplicity the coordinate flips in addition to its rotation must also be considered. Also the coordinates denoted by the eigenvectors are not unique which means that all eigenspaces must be considered to guarantee uniqueness of the coordinates of the vertices. Hence, this approach is problematic in practice.

Kunegis *et al.*, 2012 [36], suggest that since evolution or changes in a graph over time will lead to changes in its spectrum therefore an Eigen decomposition of the adjacency matrix can be used to characterise the this change. They then use the networks spectrum for link prediction and also discuss a method to reducing this link prediction problem to a 1D curve fitting problem.

In Duchenne *et al.*, 2014 [38] formulate the hypergraph matching problem as a maximization of a multilinear objective function over a tensor representing feature permutations. The tensor represents the affinity between tuples of features. A multidimensional power method is used to solve the problem and the solution is then projected onto to the closest assignment matrix. The power method utilises a tensor Eigen decomposition and is applied to point matching using some similarity measure.

2.8. Visual Methods

More recently, the authors in [39] 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.

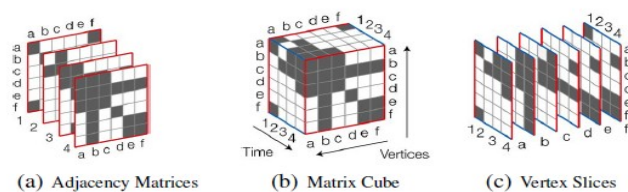


Figure 3: The matrix cube view of dynamic networks that allow for visual characterisation of network evolution over time.[39]

3. Methods

4. Results

5. Discussion

6. Evaluation, Reflections & Conclusions

7. References

- [1] A. Dickson, "Introduction to Graph Theory The Konigsberg Bridge Problem," *October*, no. October, pp. 1-15, 2006.
- [2] T. Ruya, M. Ning, L. Qianqian, and L. Yijun, "The Evolution and Application of Network Analysis Methods," *2013 IEEE Int. Conf. Syst. Man, Cybern.*, pp. 2197-2201, 2013.
- [3] S. Chen, "Graph Theory and Finance in Mathematica," *Wolfram Blog*, 2015. [Online]. Available: <http://blog.wolfram.com/2012/06/01/graph-theory-and-finance-in-mathematica/>.
- [4] A. Ma'ayan, "Introduction to Network Analysis in Systems Biology," *Sci. Signal.*, vol. 4, no. 190, p. tr5-tr5, Sep. 2011.
- [5] 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.
- [6] B. Klimt and Y. Yang, "Introducing the Enron Corpus," *Mach. Learn.*, vol. stitutep1, p. wwceasccaers2004168, 2004.
- [7] 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.
- [8] K. M. Carley, "Dynamic Network Analysis," no. September, 2001.
- [9] 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.
- [10] K. M. Carley, "Ora: Organization risk analyzer," *Science (80-.)*, no. January, p. 49, 2004.
- [11] T. A. B. Snijders, "Introduction To Dynamic Social Network Analysis," *BRIMS 2015*, no. November, 2015.
- [12] S. P. Borgatti and M. G. Everett, "A Graph-theoretic perspective on centrality," *Soc. Networks*, vol. 28, no. 4, pp. 466-484, 2006.
- [13] S. P. Borgatti, "Centrality and network flow," *Soc. Networks*, vol. 27, no. 1, pp. 55-71, 2005.
- [14] G. Costantini, S. Epskamp, D. Borsboom, M. Perugini, R. Möttus, 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.
- [15] 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.
- [16] R. Kaur and S. Singh, "A survey of data mining and social network analysis based anomaly detection techniques," *Egypt. Informatics J.*, 2015.
- [17] 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].
- [18] 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.
- [19] C. McDiarmid and F. Skerman, "Modularity in random regular graphs and lattices," *Electron. Notes Discret. Math.*, vol. 43, pp. 431-437, 2013.
- [20] F. Ashby and D. Ennis, "Similarity measures," *Scholarpedia*, vol. 2, no. 12, p. 4116, 2007.
- [21] 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.
- [22] M. Komosinski and M. Kubiak, "Quantitative measure of structural and geometric similarity of 3D morphologies," *Complexity*, vol. 16, no. 6, pp. 40-52, 2011.
- [23] D. Koutra, a Parikh, a Ramdas, and J. Xiang, "Algorithms for Graph Similarity and Subgraph Matching," 2011.
- [24] A. Rawashdeh and A. L. Ralescu, "Similarity Measure for Social Networks - A Brief Survey," 2012.
- [25] S. Soundarajan, T. Eliassi-Rad, and B. Gallagher, "A guide to selecting a network similarity method," *Sdm*, no. 1, 2014.
- [26] L. A. Zager and G. C. Verghese, "Graph similarity scoring and matching," *Appl. Math. Lett.*, vol. 21, no. 1, pp. 86-94, 2008.
- [27] "Dynamic Social Network Analysis of Al-Qaeda." .
- [28] R. Zass and A. Shashua, "Probabilistic graph and hypergraph matching," *26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, pp. 0-7, 2008.
- [29] C. Wang, L. Wang, and L. Liu, "Improving graph matching via density maximization," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 3424-3431, 2013.
- [30] 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. 3, pp. 265-298, 2004.
- [31] H. Bunke, "Recent developments in graph matching," *Pattern Recognition, 2000. Proceedings. 15th ...*, vol. 2, pp. 117-124, 2000.
- [32] 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.
- [33] 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.
- [34] H. Yaghi and H. Krim, "Probabilistic graph matching by canonical decomposition," *Image Process. 2008. ICIP 2008. 15th IEEE Int. Conf.*, pp. 2368-2371, 2008.
- [35] D. a. Spielman, "Spectral Graph Theory and its Applications," *48th Annu. IEEE Symp. Found. Comput. Sci.*, no. 1, pp. 29-38, 2007.
- [36] J. Kunegis, D. Fay, and C. Bauckhage, "Spectral evolution in dynamic networks," *Knowl. Inf. Syst.*, vol. 37, no. 1, pp. 1-36, Oct. 2013.
- [37] A. E. Brouwer and W. H. Haemers, *Spectra of Graphs*,

- vol. 1542. New York, NY: Springer New York, 2012.
- [38] 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.
- [39] 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.

