Extracting Knowledge from Complex Networks: Thesis Proposal

David McDonald
Dr. Shan He
Dr. Peter Tino

Abstract

content...

Contents

Chapter 1

Introduction

Most complex systems observed in nature can be represented as a complex network. A complex network is a highly heterogeneous object that is represented as a mathematical graph that is comprised of non-trivial and non-uniform features.

Despite being characterised as complex objects, we have observed that many real world networks, from protein interaction networks, to the internet, possess shared features that unite such seemingly disparate subject and give rise to the prevailing popularity of their study. For example, we have observed that these complex networks follow scale-free distribution of node degree (often with exponent $\gamma \in [2, 3]$) [1, 2]. Connections are preferentially made between nodes with a probability proportional to their existing degree, giving rise to the so-called preferential attachment model, a model that adheres to the old 'popularity is attractive' adage. Furthermore, complex networks also are characterised by the 'small world' phenomenon, where one would expect a small average shortest path length and a high degree of clustering [3]; and are typically very sparse with the number of edges in the same order as the number of nodes.

This work will focussed primarily upon the study of complex networks within the context of computational biology. Given the explosion of data available to us thanks to modern experimentation techniques, the need for modelling such data, has never been greater. Through the medium of complex networks, researchers wish to simplify and represent the raw data in a data-driven and scientifically justified way, such that is becomes useful for the purposes of making functional predictions and tracing the causes of observed disease phenotypes.

Already, we have seen the application of complex network models to the problems of identification of biologically relevant modules according to genetic expression profiles [4]; the multi scale detection of the function and structure of brain networks [5]; and testable protein functional predictions [6, 7]. And with the growing availability of data and drive of the network science community, models are becoming more and more complex, with the integration of multiple types of data and more sophisticated analysis techniques. It is the hope of the work proposed here to further this burgeoning field by proposing novel, integrative network models that allow for the extraction of testable hypotheses

based on the data currently available to us.

Chapter 2

Thesis Statement

The work proposed in this document shall...

2.1 Research questions

The following research questions have been selected to facilitate such an investigation:

1. Can a multi-layer network model be constructed such that the hyperbolic metric space that underpins complex network formation be uncovered for multi-layer networks?
2. What interpretable information, such as disease modules or functional predictions, can be extracted from such a model?
3. My constructing data-driven models, can we predict protein abundance and phosphorylation?

2.2 Significance

?

Chapter 3

Literature Review

In order to tackle the challenging questions posed in this work, it was necessary to undertake a thorough literature review. Network science is agnostic as to the actual data it is representing ans is instead a more general tool. Because of its inter-disciplinary nature, prolific amounts of research has been undertaken over the last decade, with researchers from all backgrounds contributing their expertise and perspective. For the purpose of brevity, the review presented in this work will only focus on the fundamental concepts that will prove to be relevant in the work proposed here. For a more general overview of network science, the reader is pointed towards [8], a book written by Albert-László Barabási, a well known figure in the field.

3.1 Community Detection

Many real-world networks contain subsets of nodes that contain a higher degree of inter-connectivity than the rest of the network [9, 6, 10]. These subsets are commonly referred to as communities. A rigorous definition for a 'community' within a network still seems to elude the scientific community [10]. However, the most popular definition among scholars is the planted l-partition model. This was popularised thanks to Girvan and Newman in their seminal work [9] and states that as long the probability of a node being connected to its group is greater than the probability of it being connected to the rest of the graph, then those groups are communities. 'Community detection' is the name given to the problem of finding the underlying community structure in a given network [9]. For example, groups of friends in a social network, functional modules in Protein Protein Interaction (PPI) networks and scientific disciplines in co-authorship networks.

But, a general community detection algorithm does not yet exist. Many existing algorithms suffer from a number of issues. To name a few: The number and scale of communities must be known a-priori, which in most real applications, is infeasible. Additionally, the relationships between communities, both

one the same level and at different ones, is lost. Identifying not only the community itself, but its position in the network as a whole, provides further insight into the often abstract interactions that comprise complex networks and so preserving this information when analysing a network is paramount. And, in some cases, the algorithms cannot deal with special cases: for example, modularity-based methods suffer from the so-called 'resolution limit' [11].

3.1.1 Cut-Based and Spectral Approaches

The flagship Kernighan-Lin algorithm [12] focused on 'cutting' the network into modules, in such a way that the number of edges cut was minimized. However, this often favoured cuts of small, peripheral subgraphs, so it was adapted into ratio cut [13], normalised cut [14] and min-max cut [15] that took the number of nodes in each resulting sub-graph into account, and thus resulted in a partition that was more balanced.

Contemporary cut-based approaches are concerned more with edges, rather than vertices and gave rise to a new measure for a good cut, called conductance:

$$\phi(S) = \frac{c_s}{\min(Vol(S), Vol(V \setminus S)} \tag{3.1}$$

with

$$c_s = |\{(u, v) : u \in S, v \notin S\}| \tag{3.2}$$

Conductance is still prolific in the literature: it has been used to detect communities in bipartite networks [16], combined with PageRank [17] and used as the basis for a greedy optimisation algorithm [18] capable of finding overlapping communities at different scales.

Spectral clustering dates back to the work of Donath and Hoffman in 1973 [19]. However, it was popularized in the early 2000s [14, 20, 21]. Spectral methods rely upon constructing 'Laplacian' matrices from the raw network data and eigen-decomposing them. Clustering the resulting eigenvectors results in clusters of the original data points. Spectral approaches have many advantages over other techniques and, as a result, they have become popular in the machine learning community for clustering on non-linear manifolds. According to [22], 'these methods do not make assumptions about the form of the clusters' and are capable of correctly identifying typically challenging clusters, such as the famous two spirals example. For community detection, they have the additional benefit of efficiency, especially if the graph adjacency matrix is sparse.

Further work includes the Markov Clustering algorithm (MCL) that simulates a diffusion process on a graph by repeatedly performing stages of expansion and inflation and only keeping the k largest elements for efficiency [23].

3.1.2 Modularity-Based Approaches

The seminal work of Girvan and Newman [9] marked a significant advance in the field by providing the first quantitative measure of a community: modularity.

The modularity of a partition of a network defined as

$$Q = \sum_{s=1}^{m} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right] \tag{3.3}$$

scores a network partition by comparing the number of links inside a given module with the expected number that would be found in a random graph of the same size and degree sequence. Here, m is the number of modules in the partition, l_s is the number of links in module s, L is the total number of links in the network and d_s is the total degree of the nodes in s. Girvan and Newman propose a hierarchical divisive algorithm that removes edges based on their 'betweenness' (the number of shortest paths from two nodes in the network that go through them) until the modularity quality function is maximized. The early work of Girvan and Newman has since been expanded upon. For example, edge clustering in favour of edge-betweenness [24], iteratively adding links to a module based on their expected increase in modularity [25], and multi-stage local optimization in the popular Louvain algorithm [26].

Other work with modularity-based quality scores include the work of Rosvall and Bergstrom with their famous Infomap algorithm [27] that translated the problem of community detection into the problem of optimally compressing the information in a graph such that the most information can be uncovered when the compression is decoded. They used simulated annealing to minimize a function that represented both compression and data loss. While slow and computationally expensive, this approach was also shown to work well with dynamic processes in his later work [28].

3.1.3 Neural Network Approaches

The deep learning community has began to explore the possibilities of using neural networks for clustering in the graph domain. Convolutional neural networks (CNNs), powerful machine learning tools that have proven very successful for challenging classification tasks that have recently been generalised to take a graph input [29]. CNNs have also been used for semi-supervised learning on graphs, where the they are capable of learning both graph structure and node features [30].

3.1.4 Hierarchical Community Detection

In many networks representing complex real-world phenomena, finding a single cover – where each node is assigned to exactly one community – does not accurately reflect the underlying community structure of the data being represented. Sometimes, nodes can belong to more than one community, and sometimes communities overlap. The first algorithm to consider overlapping communities was CFinder in 2006 [31]. Drawing from the earlier work of Palla et al. and the Clique Percolation Method (CPM) [6], CFinder considered communities as the unions of k-cliques and so rolled k-cliques across the graph to detect communities. While computationally expensive, it was able to deal with overlapping

cases, and opened the door for further study. Shen et al. proposed EAGLE in 2009 [32] that used maximal cliques, an agglomerative hierarchical structure and a modified modularity quality function that detected complex overlapping community structures.

The hierarchical nature of modularity-based clustering methods can allow them to detect communities at different scales. [10] used local optimization to maximize a fitness function with a parameter that controlled the size of communities detected. Other work includes multi-scale quality functions that can uncover hierarchical communities and produce several different partitions of a graph, the post-processing of clusters found by hierarchical methods (encoded in a dendogram) [33], and Bayesian non-negative matrix factorisation that performs 'soft-partitioning' and assigns node participation scores to modules [34].

3.1.5 Community Detection Across Multiple Layers

When more than one type of actor takes part in a complex system, a single layer network cannot accurately capture this. In these cases, one may construct a model consisting of several connected 'layers' of networks, one for each type of actor in the system. These models allow for nodes to connect to nodes in the same layer but also to nodes in other layers through special relationships. For example, one may construct a multilayer network consisting of layers of genes, connected by co-expression, proteins connected by physical interaction and inter-layer links between genes and proteins if a protein is known to related that genes expression levels. When one introduces enough types of layers, and inter-layer connections, one can consider a multilayer network to be a sort of 'network' of networks.

Community detection within these models follows from the same principles as the single layer case. In fact, many classical community detection algorithms have been generalised for application to multiplayer networks [35]. For example, [5] used a multi scale variation of the Louvain algorithm to detect hierarchical communities in functional brain networks.

These algorithms can successfully identify closely related sets of nodes, of different types, which, in the context of computational biology can be useful to predict function and missing links. Special cases of multilayer networks are the multi-slice networks where the nodes are the same for each layer but the topology of the network may differ. Examples include, temporal networks where each layer represents a different time point and networks representing many different types of interactions amongst nodes.

3.2 Active Module Identification

While community detection only considers only the structure of the network at hand, often nodes in network may be enriched with additional attributes that are not solely based upon the observed topology of the network. For example, people in a social network may be annotated with preferences such as hobbies

and interests and we would expect two people with the same interests to still somehow be similar, even if we do not observe a direct link between them in the network. Within the context of computational biology, this is perhaps even more relevant, due to the vast quantity and variety of data now available to us, and the successes of integrative models in the past. Integrative models get their name from the principle of integrating observed data (say, gene expression) with prior knowledge (often in the form of a known protein interaction network and/or previously curated functional annotations). [36] offers a summary of many the integrative approaches popular in the literature.

One of the most successful integrative approach is the identification of so-called 'active modules'. It is a relatively recent trend within the interdisciplinary fields of network science and translational medicine and aims to augment known physical interactions with observed expression levels to identify connected sub-graphs (called sub-networks for the remainder of this proposal) that are maximally differently expressed. Ideker et al. was the first to formalise this problem in [4] in 2002. Given a known PPI network G and a matrix of gene expression levels with their corresponding p-values P, we compute a z-score for each gene i in the network as:

$$z_i = \Phi^{-1}(1 - p_i) \tag{3.4}$$

and then score identified sub-networks A in an aggregated manner with

$$z_A = \frac{1}{\sqrt{k}} \sum_{i \in A} z_i \tag{3.5}$$

where a high z_A represents a biologically active sub-network. Here, Φ^{-1} is the inverse normal CDF. While the problem is related to community detection – the sub graphs must be connected – but biologically relevant modules often do not align with communities based solely on network topology.

Computing an exact solution is NP-hard [4], so the authors employ a heuristic search based on simulated annealing to search for the maximally scoring sub-graph in the network. Genetic algorithms (GAs) [37], greedy methods [38] and propagation of flow from cancer genes have since been used [39]. More recent work has employed a memetic algorithm to ensure connectedness [40]; a multi-objective optimisation process to control the trade off between biological activity and functional enrichment of the detected modules [41]; and a cooperative co-evolutionary approach [42]. Interestingly, despite the NP-hardness of the problem, [43] showed that by transforming the above problem into the well known Prize Collecting Stein Tree (PCST) problem, exact solutions can be obtained in reasonable computational time with integer programming.

3.3 Network Embedding

Several models in the literature assume the existence of an underlying metric space that controls the topology of the network. They suppose that entities

that are closer together in this space are more 'similar' and have a higher probability of being connected. These models aim to infer the geometry of these spaces and the positions of nodes within the space, such that the probability of reconstructing the observed network is maximised. This is the so-called network embedding, and is the cornerstone of the field of *network geometry*.

Network embedding is closely related to the field of manifold learning. Indeed, many classical non-linear manifold learning techniques, such as Isomap [44] and Laplacian Eigenmaps [45], must first construct nearest neighbour graphs based on dissimilarities between samples before dimensionality reduction takes place. Many of these techniques are directly applicable to embedding of complex networks by simply omitting the graph construction step.

An interesting and popular embedding paradigm in the literature comes from natural language processing. In particular, the Skipgram model and the Word2Vec algorithm that aims to vectorise words and phrases in a semantic space such that similar words are mapped close together [46, 47]. The principle idea is, given a corpus of words and a particular sentence, generate a 'context' for each input word with the aim of maximising the likelihood of observing context words in the embedding space, given the input word. Similarities are measures by dot products and accordingly, observation probabilities are computed using a multilayer perception with a linear hidden layer and softmax output. Through the use of sub-sampling and negative sampling (replacing softmax with sigmoid), training can be made very efficient and the resulting embeddings can be obtained from the activation of the hidden units. This idea naturally extends to networks, where sentences are replaced by 'neighbourhood graphs' generated from random walks. Furthermore, the shallow architecture of the Skipgram model has been replaced with multiple non-linear layers to learn the highly non-linear relationships between nodes [48, 49]. By introducing additional parameters into the random walk to control a breadth vs. depth first neighbourhood search, [50] were able to identify neighbourhoods of nodes with high *homophily* and high structural similarity.

3.3.1 Embedding to a Hyperbolic Metric Space

An emerging popular belief in the literature is that the underlying metric space of most complex networks is in fact hyperbolic. Nodes in real world networks often form a *taxonomy*, where nodes are grouped hierarchically into groups in an approximate tree structure. Hyperbolic spaces can be viewed as continuous representations of this tree structure and so models that embed networks into hyperbolic space have proven to be increasingly popular in the literature [51, 52]. In fact, this assumption has already had proven success in the task of greedy forwarding of information packets where nodes use only the hyperbolic coordinates of their neighbours to ensure packets reach their intended destination [53].

The most popular of all these models is the Popularity-Similarity (or PS) model [54]. This model extends the "popularity is attractive" aphorism of preferential attachment to include node similarity as a further dimension of attach-

ment. Nodes like to connect to popular nodes but also similar ones. The PS model sustains that the clustering and hierarchy observed in real world networks is the result of this principle [55]. This model has a simple interpretation in two dimensional hyperbolic space, where nodes are placed on a hyperbolic disk, with radial coordinates representing popularity and angular coordinates representing similarity. Then the hyperbolic distance between two nodes $\mathbf{x}_1 = (r_1, \theta_1)$ and $\mathbf{x}_2 = (r_2, \theta_2)$, given by the hyperbolic law of cosines:

$$d(\mathbf{x}_1, \mathbf{x}_2) = \operatorname{arccosh}(\cosh(r1)\cosh(r2) - \sinh(r1)\sinh(r2)\cos(\Delta\theta) \tag{3.6}$$

$$\Delta\theta = \pi - |\pi - |\theta_1 - \theta_2|| \tag{3.7}$$

controls their connection probabilities. Nodes with short hyperbolic distances show a higher probability of being connected.

Maximum likelihood (ML) was used in [54] to search the space of all PS models with similar structural properties as the observed network, to find the one that fit it best. This was extended by the authors in [56, 57]. Due to the computationally demanding task of maximum likelihood estimation, often heuristic methods are used. For example, [55] used Laplacian Eigenmaps to efficiently estimate the angular coordinates of nodes in the PS model. The authors then combined both approaches to leverage the performance of ML estimation against the efficiency of heuristic search with a user controlled parameter in [58]. Additionally, [59] propose the use of classical manifold learning techniques in the PS model setting with a framework that they call *coalescent embedding*.

3.3.2 Embedding with Node Dynamics

Little work has been done to embed networks, accounting for both topology and node attributes, however recent years has shown it to become more prolific. [60] embed into a vector space based on the statistics of attributes and pairs of attributes, [61] draw from the well known fields of manifold learning and multi-view learning to align the projections based on topology and attributes and [62] use deep learning. In [63], the authors generalised convolutional neural networks from regular pixel lattices to arbitrary graphs. It is worth noting that by transforming an unweighted graph into a so-called 'flow graph' [64] by weighting links by node expression, many embedding techniques that are applicable to weighted graphs can be applied to unweighted graphs with node attributes. However, it is not clear how to do this if the graphs are already weighted, or nodes are annotated with discrete or multiple attributes.

Chapter 4

Proposed Work

After a thorough literature review, it became apparent that there are a number of gaps worthy of further investigation.

4.1 Module Identification

First and foremost, the identification of modules has proven to be an important step in extracting knowledge from a complex network. It is clear from the literature that topological modules (communities) rarely align with complete functional modules, such as pathways [42]. Because of this, the addition of prior knowledge is essential to guide the module search to look beyond topology alone. Similarly, the identification of active modules must be guided by both node attributes (gene expressions) and network topology (identified subnetworks must be connected) [4]. Furthermore, modules in biological networks have been shown to have a highly hierarchical structure – where large modules of more general function are composed of many smaller ones that perform more specific tasks. As such, any model for meaningful biological module recovery must take hierarchy into account. Also, biological systems are dynamic – in that their activity, and maybe even topology changes over time – and composed of many types of entities and interactions. Because of this, network models must be scalable from single to multiple layers in tractable time. Finally, due to experimental error and measurement difficulties often data is incomplete and so any model must be robust enough to handle this. All of these factors combined make the problem of identifying the modular structure of biological systems a complex optimisation and combinatorial problem. This work intends to overcome some of the weaknesses of other techniques – such as poor scalability or user-defined parameters to control the number of modules or module scale – by developing models that allow for the extraction of modules in a more data-driven way.

The first such model will use the topology preserving property of the self-organising map to construct a map of communities based on an observed net-

work. The model will use the Growing Hierarchical Self-Organising Map (GHSOM) [65] to grow maps organically with no need to specify map size a priori. Furthermore, neurons in the map may be selected for expansion, resulting in new maps of higher granularity for those data points. This results in a set of maps with hierarchical structure that reflects the underlying hierarchical community structure of the data.

This model has a number of drawbacks, in that it is solely topology based, and requires some network embedding to a Euclidean space or adaption of the algorithm to work directly on graphs, as in [66]. The map structure may also impose too strong an emphasis on module connection and so the Growing Hierarchical Neural Gas [67] may be a more suitable choice, as it only connections between neurons are allowed also form in a data-driven way. It is also unclear how to scale this module to multiple layers without utilising some multi-view learning techniques to align the embeddings of the layers into the Euclidean space. Also, despite the appealing nature of not needing to know the number of modules or structure a priori, GHSOM is controlled by a number of parameters that can be difficult to tune.

4.2 Hyperbolic Network Embedding

Network embedding can be considered a preprocessing step towards knowledge extraction, such as module identification. The previous section mentioned an embedding step before module identification could take place, for example. Embedding allows for the transformation of problems: from module to detection to clustering. But, selecting a metric space and embedding method is a very non-trivial problem.

The recent trend towards models that support an underlying hyperbolic metric space is very appealing, as results are extremely positive, and the space is a generalisation of the hierarchical structure that we know actually exists in complex networks. However, there are a number of questions as yet unanswered that this work shall focus on. Firstly, how to embed networks with node attributes into hyperbolic space, such that active modules can be identifies as clusters. While single continuous node attributes can be combined with into the topology of an unweighted graph via construction of flow graphs [64], some of the current hyperbolic embedding methods (for examples LABNE [55]) is unsuitable for weighted graphs. Furthermore, flow graphs cannot be used if there is more than one node attribute. Another challenge is embedding of multilayer networks.

This work will first attempt to tackle both of these problems for hyperbolic network embedding with the application of multi-view learning techniques. We will consider each layer as a separate view and align the separate embeddings onto the same space using, for example, kernel canonical correlation analysis (KCCA) [68]. In the case of networks with node attributes, we shall first construct an 'attribute network' by connecting nodes with similar attributes, for example with KNN. Another technique that shall be adapted for the hyperbolic

space is the recently proposed extension of convolutional neural networks to arbitrary graphs [63]. Additional ideas for future work include extending the currently used 2D hyperbolic plane to three dimensions to allow for simultaneous embedding of all layers. However, very little thought has been put into this currently.

4.3 Protein Abundance Prediction

Protein function almost always determines the biological function of a cell. As a result, proteomics (the study of proteins) is perhaps the most well studied area of computational biology. However, there are a number of challenges that face the proteomics research community today. Firstly, protein datasets are often incomplete, due to experimental errors and other challenges. As such, an method to effectively impute missing values would enable the community to access data that would otherwise be ignored in analyses. Furthermore, protein data is expensive to generate and not as readily available to the scientific community as Copy-Number Alterations (CNA) and mRNA data. It is well known that mRNA is used to transfer information to proteins, however it has been shown that RNA expression alone is weakly predictive of protein levels [69]. We would like to augment this analyses with the integration of CNA profiles – which have been shown to affect protein abundances [7] – for more accurate protein abundance prediction. Finally, by further integration of known proteomics profiles, we would like to predict the phosphorylation levels of proteins, as this has been shown to be one of the key so-called Post-translational Modifications (PTMs) that affect protein, and therefore cell, function.

These challenges form the basis for the NCI-CPTAC DREAM Proteogenomics Challenge that this author, along with a distinguished team of researchers, shall participate in.

Tackling the imputation challenge shall begin by performing a servery and comparison against ground truth of many common imputation techniques found in the literature. For example, sample means, matrix factorization, spectral regularization [70], Multiple Imputation by Chained Equations (MICE), and an autoencoder with an adapted objective as in [71]. The best algorithm shall be adapted to give improved performance.

Protein abundance and phosphorylation levels shall be tacked in a three stage way. First, by the application of simple machine learning techniques, such as regression models. Then, by constructing network models. Finally, by augmenting our models with the inclusion of domain knowledge than can be extracted from the given data, such as ribosome number and binding site information, provided by an expert in the field.

Chapter 5

Preliminary Results

SOM paper – results in question to be honest
figures here

Chapter 6

Work Plan / Timeline

SOM for topology preserving hierarchical community detection - to be updated
Hyperbolic Node Embedding with Dynamics
DREAM proteogenomics challenge
Active Modules

Chapter 7

Implications of Research

READ PAPERS FOR JUSTIFICATION OF VARIOUS AREAS

Extraction of knowledge from vast volume of data produced from experiments -visualisation -removal of unnecessary clutter (only modules vs whole network etc.) -easier for biologist interpretation

Understanding of the formation of complex networks -link prediction (Y2H experimentation is expensive) -function prediction/annotation

Identification of biologically relevant modules - multi-layer: genes/proteins/metabolites - understanding of functional/topological/dynamical modules

Data imputation/protein abundance/phosphorylation prediction: Proteomic profiling data from mass spectrometry based experiments often contain a large number of missing values due to the dynamic nature of the mass spectrometry instruments. RNA/DNA cheaper to generate and more widely available than proteomic data RNA alone is only weakly predictive [69] Protein abundance / cellular activity determined by many post-translational modifications (PTMs) -enrichment testing for phosphorylation is expensive -improve biomarker development

Bibliography

[1] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[2] Albert-László Barabási. Scale-free networks: a decade and beyond. *science*, 325(5939):412–413, 2009.

[3] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442, 1998.

[4] Trey Ideker, Owen Ozier, Benno Schwikowski, and Andrew F Siegel. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, 18(suppl_1):S233–S240, 2002.

[5] Arian Ashourvan, Qawi K Telesford, Timothy Verstynen, Jean M Vettel, and Danielle S Bassett. Multi-scale detection of hierarchical community architecture in structural and functional brain networks. *arXiv preprint arXiv:1704.05826*, 2017.

[6] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.

[7] Hui Zhang, Tao Liu, Zhen Zhang, Samuel H Payne, Bai Zhang, Jason E McDermott, Jian-Ying Zhou, Vladislav A Petyuk, Li Chen, Debjit Ray, et al. Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell*, 166(3):755–765, 2016.

[8] Albert-László Barabási. Linked: How everything is connected to everything else and what it means. *Plume Editors*, 2002.

[9] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[10] Andrea Lancichinetti, Santo Fortunato, and János Kertész. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 11(3):033015, 2009.

[11] Santo Fortunato and Marc Barthelemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41, 2007.

[12] Brian W Kernighan and Shen Lin. An efficient heuristic procedure for partitioning graphs. *Bell system technical journal*, 49(2):291–307, 1970.

[13] Y-C Wei and C-K Cheng. Ratio cut partitioning for hierarchical designs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 10(7):911–921, 1991.

[14] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

[15] Chris HQ Ding, Xiaofeng He, Hongyuan Zha, Ming Gu, and Horst D Simon. A min-max cut algorithm for graph partitioning and data clustering. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 107–114. IEEE, 2001.

[16] Michael J Barber. Modularity and community detection in bipartite networks. *Physical Review E*, 76(6):066102, 2007.

[17] Reid Andersen, Fan Chung, and Kevin Lang. Local graph partitioning using pagerank vectors. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 475–486. IEEE, 2006.

[18] Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: a comparative analysis. *Physical review E*, 80(5):056117, 2009.

[19] William E Donath and Alan J Hoffman. Lower bounds for the partitioning of graphs. *IBM Journal of Research and Development*, 17(5):420–425, 1973.

[20] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.

[21] Chris Ding. A tutorial on spectral clustering. In *Talk presented at ICML.*, 2004.

[22] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[23] Stijn Marinus Van Dongen. *Graph clustering by flow simulation.* PhD thesis, 2001.

[24] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2658–2663, 2004.

[25] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.

[26] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[27] Martin Rosvall and Carl T Bergstrom. An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences*, 104(18):7327–7331, 2007.

[28] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.

[29] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *arXiv preprint arXiv:1606.09375*, 2016.

[30] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[31] Balázs Adamcsek, Gergely Palla, Illés J Farkas, Imre Derényi, and Tamás Vicsek. Cfinder: locating cliques and overlapping modules in biological networks. *Bioinformatics*, 22(8):1021–1023, 2006.

[32] Huawei Shen, Xueqi Cheng, Kai Cai, and Mao-Bin Hu. Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications*, 388(8):1706–1712, 2009.

[33] Pascal Pons and Matthieu Latapy. Post-processing hierarchical community structures: Quality improvements and multi-scale view. *Theoretical Computer Science*, 412(8):892–900, 2011.

[34] Ioannis Psorakis, Stephen Roberts, Mark Ebden, and Ben Sheldon. Overlapping community detection using bayesian non-negative matrix factorization. *Physical Review E*, 83(6):066114, 2011.

[35] Peter J Mucha, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *science*, 328(5980):876–878, 2010.

[36] Koyel Mitra, Anne-Ruxandra Carvunis, Sanath Kumar Ramesh, and Trey Ideker. Integrative approaches for finding modular structure in biological networks. *Nature reviews. Genetics*, 14(10):719, 2013.

[37] Martin Klammer, Klaus Godl, Andreas Tebbe, and Christoph Schaab. Identifying differentially regulated subnetworks from phosphoproteomic data. *BMC Bioinformatics*, 11(1):351, Jun 2010.

[38] Şerban Nacu, Rebecca Critchley-Thorne, Peter Lee, and Susan Holmes. Gene expression network analysis and applications to immunology. *Bioinformatics*, 23(7):850–858, 2007.

[39] Fabio Vandin, Eli Upfal, and Benjamin J Raphael. Algorithms for detecting significantly mutated pathways in cancer. *Journal of Computational Biology*, 18(3):507–522, 2011.

[40] Dong Li, Zhisong Pan, Guyu Hu, Zexuan Zhu, and Shan He. Active module identification in intracellular networks using a memetic algorithm with a new binary decoding scheme. *BMC genomics*, 18(2):209, 2017.

[41] Weiqi Chen, Jing Liu, and Shan He. Prior knowledge guided active modules identification: an integrated multi-objective approach. *BMC systems biology*, 11(2):8, 2017.

[42] Shan He, Guanbo Jia, Zexuan Zhu, Daniel A Tennant, Qiang Huang, Ke Tang, Jing Liu, Mirco Musolesi, John K Heath, and Xin Yao. Cooperative co-evolutionary module identification with application to cancer disease module discovery. *IEEE Transactions on Evolutionary Computation*, 20(6):874–891, 2016.

[43] Marcus T Dittrich, Gunnar W Klau, Andreas Rosenwald, Thomas Dandekar, and Tobias Müller. Identifying functional modules in protein–protein interaction networks: an integrated exact approach. *Bioinformatics*, 24(13):i223–i231, 2008.

[44] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

[45] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in neural information processing systems*, pages 585–591, 2002.

[46] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[47] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[48] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1067–1077. International World Wide Web Conferences Steering Committee, 2015.

[49] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710. ACM, 2014.

[50] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864. ACM, 2016.

[51] Dmitri Krioukov, Fragkiskos Papadopoulos, Amin Vahdat, and Marián Boguñá. Curvature and temperature of complex networks. *Physical Review E*, 80(3):035101, 2009.

[52] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguná. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, 2010.

[53] Fragkiskos Papadopoulos, Dmitri Krioukov, Marián Boguñá, and Amin Vahdat. Greedy forwarding in dynamic scale-free networks embedded in hyperbolic metric spaces. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.

[54] Fragkiskos Papadopoulos, Maksim Kitsak, M Serrano, Marián Boguná, and Dmitri Krioukov. Popularity versus similarity in growing networks. *arXiv preprint arXiv:1106.0286*, 2011.

[55] Gregorio Alanis-Lobato, Pablo Mier, and Miguel A Andrade-Navarro. Efficient embedding of complex networks to hyperbolic space via their laplacian. *Scientific Reports*, 6, 2016.

[56] Fragkiskos Papadopoulos, Constantinos Psomas, and Dmitri Krioukov. Network mapping by replaying hyperbolic growth. *IEEE/ACM Transactions on Networking (TON)*, 23(1):198–211, 2015.

[57] Fragkiskos Papadopoulos, Rodrigo Aldecoa, and Dmitri Krioukov. Network geometry inference using common neighbors. *Physical Review E*, 92(2):022807, 2015.

[58] Gregorio Alanis-Lobato, Pablo Mier, and Miguel A Andrade-Navarro. Manifold learning and maximum likelihood estimation for hyperbolic network embedding. *Applied Network Science*, 1(1):10, 2016.

[59] Josephine Maria Thomas, Alessandro Muscoloni, Sara Ciucci, Ginestra Bianconi, and Carlo Vittorio Cannistraci. Machine learning meets network science: dimensionality reduction for fast and efficient embedding of networks in the hyperbolic space. *arXiv preprint arXiv:1602.06522*, 2016.

[60] Jaume Gibert, Ernest Valveny, and Horst Bunke. Graph embedding in vector spaces by node attribute statistics. *Pattern Recognition*, 45(9):3072–3083, 2012.

[61] Jundong Li, Harsh Dani, Xia Hu, Jiliang Tang, Yi Chang, and Huan Liu. Attributed network embedding for learning in a dynamic environment. *arXiv preprint arXiv:1706.01860*, 2017.

[62] Lizi Liao, Xiangnan He, Hanwang Zhang, and Tat-Seng Chua. Attributed social network embedding. *arXiv preprint arXiv:1705.04969*, 2017.

[63] Mathias Niepert, Mohamed Ahmed, and Konstantin Kutzkov. Learning convolutional neural networks for graphs. In *International Conference on Machine Learning*, pages 2014–2023, 2016.

[64] R. Lambiotte, R. Sinatra, J-C Delvenne, T. S. Evans, M. Barahona, and V. Latora. Flow graphs: Interweaving dynamics and structure. *PHYSICAL REVIEW E*, 84, 2011.

[65] Michael Dittenbach, Dieter Merkl, and Andreas Rauber. The growing hierarchical self-organizing map. In *IJCNN (6)*, pages 15–19, 2000.

[66] Takeshi Yamakawa, Keiichi Horio, and Masaharu Hoshino. Self-organizing map with input data represented as graph. In *International Conference on Neural Information Processing*, pages 907–914. Springer, 2006.

[67] Esteban J Palomo and Ezequiel López-Rubio. The growing hierarchical neural gas self-organizing neural network. *IEEE transactions on neural networks and learning systems*, 2016.

[68] David R Hardoon, Sandor Szedmak, and John Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12):2639–2664, 2004.

[69] Christine Vogel and Edward M Marcotte. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature reviews. Genetics*, 13(4):227, 2012.

[70] Rahul Mazumder, Trevor Hastie, and Robert Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *Journal of machine learning research*, 11(Aug):2287–2322, 2010.

[71] Brett K Beaulieu-Jones, Jason H Moore, THE POOLED RESOURCE OPEN-ACCESS ALS, and CLINICAL TRIALS CONSORTIUM. Missing data imputation in the electronic health record using deeply learned autoencoders. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, volume 22, page 207. NIH Public Access, 2016.