Thermodynamic Motif Analysis for Directed Stock Market Networks

Dongdong Chen[a], Xingchen Guo[b], Jianjia Wang[c], Jiatong Liu[a], Zhihong Zhang*[a], Edwin R. Hancock[d]

[a]*Xiamen University, Xiamen, China*
[b]*Xian Jiaotong University, Xian, China*
[c]*Shanghai University, Shanghai, China*
[d]*University of York, York, UK*

Abstract

In this paper, we present a novel thermodynamically based analysis method for directed networks, and in particular for time-evolving networks in the finance domain. Based on an analogy with a dilute gas in statistical mechanics, we develop a partition function for a network composed of directed motifs. The method relies on the decomposition of directed networks into a series of frequently occurring graphlets, or motifs. According to the connection between a directed network and the dilute gas, the network motifs have the same topological structure as the low-order interactions between particles in the gas. This means that we can use the so-called cluster expansion from statistical mechanics to develop a partition function for the motif decomposition. In prior work, we have reported a detailed analysis of the cluster expansion for the case of undirected graphs, and showed how the resulting motif entropy can be used to analyse time evolving networks [1]. In this paper we extend this work to the case of directed graphs to compute thermodynamic quantities including energy, entropy and temperature for the directed network. The three thermodynamic quantities constitute the thermodynamic framework for the analysis of directed network evolution. We apply our thermodynamic framework to the financial and biological domains to represent real world complex systems as time-varying

*Corresponding author: zhihong@xmu.edu.cn (Zhihong Zhang)

directed networks. Experimental results successfully demonstrate the effectiveness of the thermodynamic framework in representing the evolution of directed network structure and anomalous event detection.
Keywords: Cluster Expansion, Motif, Directed Network Entropy

1. Introduction

Research on the learning and inference of network representation has become a central topic in the field of pattern recognition and machine learning. In particular, the use of network representations as succinct characterizations of time varying complex sysems [2] has made possible the understanding of otherwise intractable data from biological, social and technological domains. The network representation allows the complex patterns of interactions between the entities in such complex systems to be modelled in terms of basic distributions of its connection characteristics. Such systems can be well captured by the statistics of the underlying structure of the network [3]. However, these basic representations tend to ignore the importance of local multiplexing and similar proximity structures in networks. Recently, recurring patterns, termed network motifs, have been proved to be useful in providing the basic building blocks for different subgraphs which perform specific functional roles in a larger network structure [4, 5].

Motifs have been proved to provide an efficient way to uncover the structural design principles of a system represented by a complex network. Examples include positive and negative autoregulation [6, 7], positive and negative cascades [8, 9, 10], positive and negative feedback loops [11], feedforward loops (FFLs) [12, 13], single input modules [14], and combinations of these too [15, 16]. With the further study of network motifs, many methods for motif detection and counting have been developed [17], but how to combine the structural characteristics of the motifs with their statistical characteristics to represent the entire complex network remains an open problem. This task requires an understanding of the basic structural elements constituting the motifs and the processes

which give rise to them from a microscope point of view [18]. To embark on this type of analysis, tools from statistical mechanics provide a convenient route to the characterization of network structure.

When a thermodynamic system is modelled using statistical mechanics, then the partition function which captures the probabilities of the different microstates of the system can be used to compute thermodynamic properties such as energy, entropy and temperature [19]. Statistical mechanics can be combined with graph theory to provide a practical framework for complex networks, especially highly structured and time-evolving networks [20, 21]. For example, Delvenne and Libert [22] compute the stationary distribution for large networks through the Ruelle-Bowens random walk [23], and deal with disconnected networks with a centrality measurement which they refer to as Entropy Rank. Estrada [24] also interprets the subgraph centrality as a partition function of a network, and has defined the thermodynamic quantities on the basis of spectral graph theory.

In prior work we have provided the building blocks for the study reported in this paper. Namely, we have used the cluster expansion from statistical mechanics to perform a graphlet or motif decomposition of network structure for undirected graphs [1]. Both motifs and graphlets are subgraphs of a larger network. Here we decompose the original graph according to the structure by the subgraphs that we have predefined. This prior work has mainly investigated the mapping from clusters in the gas model to undirected network motifs, and derived the partition function for undirected networks based on the classical cluster expansion. The corresponding entropy of an undirected graph can be derived from the partition function and used to effectively represent the evolution of network structure. This has verified the validity of the partition function for undirected networks. However, this work is therefore limited to the study of undirected graphs, and was restricted to deriving a simple expression for undirected network entropy. In this paper, we resume the analysis and further consider the more detailed microscopic structure of the network, especially that resulting from directed edges. We use this to develop a thermodynamic frame-

work for computing the internal evolutionary characteristics of the directed network in terms of its motif composition.

2. Related Work

2.1. Network Motifs

There is a substantial literature on analyzing network motifs, starting from a subgraph description of the frequently occurring patterns within a network structure. Motifs are recurring patterns that can be used in the representation of more complex structure [4, 25]. They reflect not only the structural properties of a network but can also capture its functional properties. Motifs can also be regarded as the fundamental building blocks of complex networks since identical network motifs exist in diverse fields such as biology and sociology [26]. For example, in the biological domain, network motifs are implicated in both signaling [27] and neuronal activities [28], and also account for the integration of transcriptional regulation and protein-protein interactions [29].

Due to the deep insights provided by the motif structure of networks and the direct links between motif structure and the function of specific real-world networks, many algorithms to detect and enumerate the frequency of network motifs have been proposed [30, 31]. These algorithms include exact counting [32], sampling [33], and pattern growth methods [34], and are developed under a number of different paradigms. Network-motif patterns can also be identified when the nodes and edges in a network are annotated with quantitative features. The expected number of appearances of a motif can be determined using a Null-model [30], which is an ensemble of random networks with some properties in common with the original network.

2.2. Cluster Expansion

In statistical mechanics, the cluster expansion is usually a power series expansion of the partition function. Each order of expansion describes the primitive patterns of interaction in a system with a large number of particles [35]

in terms of smaller units of interaction. Mayer and his collaborators [36] were the first to carry out a systematic study of alternative expansions, in the case of real gases obeying classical statistics. Kahn and Uhlenbeck [37] generalize the cluster expansion to gases obeying quantum statistics. Lee and Yang [38] explore the application of these ideas to the real world.

The cluster expansion has become a standard tool in the analysis of solid-state structures [39, 40]. Examples include the theory of two-dimensional solids, which are a kind of regular network structure [41]. Cao and Li [42] further demonstrate how the cluster expansion assists in building structureproperty relationships and hence predict the functional structures occurring in a two-dimensional square lattice.

The remainder of the paper is structured as follows. In Sec 3, we first show how the partition function in the classical gas system is linked to the directed networks based on the description of graph motifs. With this to hand, we then provide a detailed account of the development of several thermodynamic variables of directed networks, i.e., the motif entropy, average motif energy, and temperature. In Sec 4, we apply the resulting thermodynamic characterization to several real-world time-varying directed networks, including the drosophila melanogaster gene expression data, the New York Stock Exchange (NYSE) data, and the financial networks in the past 10 years. Finally, in Sec 5, we summarize the main contributions of this paper and also suggest a few research directions for the future.

3. Thermodynamic Framework for Directed Networks

In this section, we provide a detailed description of how we construct our thermodynamic framework for networks using graph motifs, especially to compute the three thermodynamic quantities, namely the energy, the entropy, and the temperature.

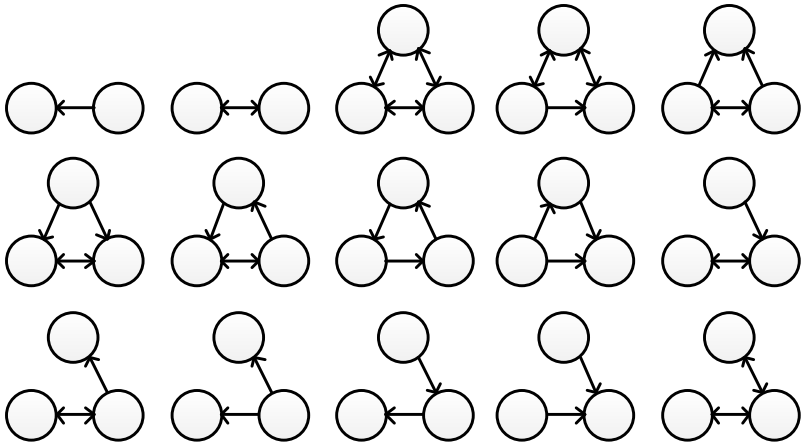


Figure 1: Typical motifs for a directed graph. The arrows indicate the direction of the edge. The edges are both unidirectional and bidirectional. Unidirectional edges have a single arrow. Bi-directional edges have two opposed arrows.

3.1. Initial considerations

Let $G =< V, E >$ represent a graph where V is a set of vertices, and $E \subseteq V \times V$ is the set of edges. Further, let $N = |V|$ be the total number of nodes in the graph G. We can decompose the graph into differently sized motifs, which are the representative subgraphs that frequently appear in a graph. Suppose that the individual motif index is ν and the ν-th motif has node-set $\mathcal{L}_\nu$ and edge-set $\mathcal{M}_\nu$. Further let l_ν be the number of nodes in the ν-th motif and n_ν be the frequency of occurrence of the ν-th motif. We will exploit an analogy with the classical non-ideal gas system from physics. This system is composed of N particles which occupy a volume Vol_{gas}. For such a gas we can represent interactions between particles using the connected graph composed of l particles, which is referred to as the l-cluster. The number of l-clusters in the classical gas system is m_l, and thus we have the constraint that

$$\sum_l l m_l = N.$$

3.2. Directed Motif Detection

Later on in this paper, we implement our method by detecting 15 distinct motifs for each graph. The structure of the motifs is shown in Figure 1. The detailed steps of the motif detection process are listed in Algorithm 1.

3.3. Partition Function

We begin by briefly reviewing the structure of the partition function Z based on the cluster expansion for the classical gas system together with the computation of necessary cluster expansion coefficients b_l. These coefficients account for the motif-based interactiosn and are the key element of our previously published work on undirected motifs [1].

For undirected graphs, the partition function for the classical gas system using the notation described above is

$$Z = \sum_{\nu \in \chi} \prod_{l=1}^{|\mathcal{L}_\nu|} \frac{1}{m_l!} \left\{ \frac{Vol_{gas}}{\lambda^3} b_l \right\}^{m_l} \tag{1}$$

$$b_l = \frac{1}{l! \lambda^{3l-3} Vol_{gas}} \int ... \int \sum \prod_{i<j\leq l} f_{i,j} d^3\vec{r}_1 d^3\vec{r}_2 ... d^3\vec{r}_l \tag{2}$$

where $\nu \in \chi$ is an index that represents the different kinds of lower-order interactions which give rise to clusters in the particle system, χ is the set of such interactions or motifs, and l is the number of particles constituting the system. The product $\prod_{l=1}^{|\mathcal{L}_\nu|}$ is over all particles in the different clusters. The scale paramerer λ is the average de Broglie wavelength of the particles in the gas. For the cluster of particles indexed l, m_l is the frequency of occurrence and b_l is the expansion coefficient which is given by a classical cluster integral for the l-cluster. Finally, $f_{i,j}$ is the interaction strength (or Mayer function) between the nodes (or particles) i and j.

To evaluate the motif partition function for directed networks, on the other hand, we again map the directed network motifs to the classical cluster expansion. The detailed mapping process can be seen in Table 1. The particles in the gas system are nodes of the network, and interactions between particles

Algorithm 1: Detecting Directed Graph Motifs

Input: graph $G = (V, E)$, type of motifs m

Output: A set of motif number for each type of motif $\{n_1, ..., n_\nu, ..., n_m\}$

initialization, Adjacency matrix of the graph A, Number of nodes in the the Graph N ;

Re-sort each row and column of adjacency matrix according to the degree of the corresponding node from small to large A;

If $Node_i$ connected to $Node_j$, $Undirectional(Node_i, Node_j)$=1

If $Node_i$ bidirected to $Node_j$, $Bidirectional(Node_i, Node_j)$=1;

Compute Two Node motif;

$n_\nu = 0$, a = copy.copy(A);

for *i in range(N)* **do**

  for *j in range(i, N)* **do**

    if *Undirectional(i, j)* **then**

      $n_1 += 1$;

    end

    if *Bidirectional(i, j)* **then**

      $n_2 += 1$;

    end

  end

end

Compute Three Node motif;

$n_\nu = 0$, a = copy.copy(A);

for *i in range(N)* **do**

  for *j in range(i, N)* **do**

    for *k in range(k, N)* **do**

      Judge the connection mode of the three nodes;

      if *Node i, j, k construct the ν-th motif* **then**

        $n_\nu += 1$;

      end

    end

  end

end

return n for the number of ν-th motif in different structure ;

Table 1: Comparison between thermodynamic gas system and directed network.

Gas System	Directed Network
cluster	motif
number of particles in gas system N	number of nodes in the network N
multiple volume for the gas system Vol_{gas}	single scalar radial variable r
number of particles in an l-cluster l	number of nodes in the ν-th motif l_ν
frequency of occurrence of the l-cluster m_l	frequency of occurrence of the ν-th motif n_ν
the classical cluster integral for l-cluster b_l	the configuration integral for the ν-th motif q_ν

correspond to edges. In terms of the computational elements of the mapping, we treat the motifs as the interactions represented by the expansion coefficients appearing in the partition function. We simplify the partition function to the one-dimensional case by replacing the multiple volume integrals by a single scalar radial variable r and by ignoring the dependence on λ which is a constant related to the physical properties of particles. We let n_ν be the number of motifs of type ν analogously to the number of l-clusters m_l in the classical cluster expansion. The corresponding motif configuration integral q_ν is the configuration integral for the ν^{th} motif, and this plays a similar role to b_l in the cluster expansion. The partition function for the network can be written as a sum over motifs ν as

$$Z = \sum_{\nu \in \chi} \prod_{n_\nu=1}^{|\mathcal{L}_\nu|} \frac{1}{n_\nu!} \Big\{ r q_\nu \Big\}^{n_\nu} = \sum_{\nu \in \chi} z_\nu \tag{3}$$

i.e. the sum over all possible motifs for the N particles.

We use differently sized motifs to decompose the network, and according to this decomposition the network can be considered as consisting of two different

components in terms of the interactions between particles or nodes. The first component is an interaction term governed by the connected component of a specific network motif. The second component is a non-interaction part, governed by the set of independent or disjoint nodes. Thus the number of these nodes is $N - l_\nu n_\nu$. The product over the integrals of all possible graphs becomes the product of the integral over the connected interaction components together with an disjoint non-interaction part. As a result we can write the partition function for motif ν when $N - l_\nu n_\nu \geq 0$

$$z_\nu = \frac{1}{n_\nu!}(rq_\nu)^{n_\nu}\frac{1}{(N - l_\nu n_\nu)!}(rq_0)^{N-l_\nu n_\nu} \tag{4}$$

where q_ν is the configuration integral for the ν^{th} motif which plays a similar role to b_l in the cluster expansion, and l_ν is the number of nodes in the motif. Here q_0 is the configuration integral of a single node which we set to unity.

According to classical cluster expansion,the expansion co-efficient b_l represents the configuration integral of l-cluster in Eq. 2 , and q_ν as the configuration integral of the ν-th motif in the network. In this setting the quantity analogous to b_l defined above is

$$q_\nu = \frac{1}{l_\nu! r}\int ... \int \sum \prod_{i<j\leq l} f_{i,j} d^3\vec{r}_1 d^3\vec{r}_2 ... d^3\vec{r}_l = \frac{1}{l_\nu! r}\zeta_\nu \tag{5}$$

where ζ_ν is the configuration integral resulting from product of integrals over all edges connecting nodes in the ν-th motif.

3.4. Motif Energy and Entropy

The average energy of the network can be expressed in terms of the Hamiltonian operator and the partition function as

$$U = k\beta^2[\frac{\partial}{\partial T}\ln Z] = \frac{\partial}{\partial \beta}\ln Z \tag{6}$$

Moreover, the thermodynamic entropy S is obtained by

$$S = \ln Z - \beta\frac{\partial \ln Z}{\partial \beta} \tag{7}$$

For directed graphs, we capture the directionality of edges connecting the nodes of a motif using the number of unidirectional edges $d_{u,\nu}$ and the number of bidirectional edges $d_{b,\nu}$ in the ν-th motif. Thus the number of disjoint nodes without connecting edges is $l_\nu - d_{u,\nu} - d_{b,\nu}$. The directed motif configurational integral can be obtained by decomposing the overall configuration integral into contributions originating from nodes connected by unidirectional edges, bi-directional edges and those that are disjoint.

$$\zeta_\nu = \epsilon_0^{l_\nu - d_{u,\nu} - d_{b,\nu}} (\epsilon_u)^{d_{u,\nu}} (\epsilon_b)^{d_{b,\nu}} \tag{8}$$

where ϵ_u and ϵ_b are the corresponding edge configurational integrals obtained using the Mayer function and given by

$$\begin{aligned} \epsilon_u &= \int_0^\infty (e^{-\beta v(r)} - 1) dr \\ \epsilon_b &= \int_0^\infty (e^{-\beta v(r)} - 1)^2 dr \end{aligned} \tag{9}$$

where r is the separation betweeen particles in the gas (or the "length" of the corresponding edge) Here we model the inter-particle repulsive potential vr) using the Lennard-Jones potential

$$v(r) = 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^6]$$

from chemical-physics. To make numerical computations of the configuration integrals, we use Simpson's Method [43]

$$\begin{aligned} \epsilon_u &= e^\beta \sum_{r=r_{min}}^{r_{max}} e^{-v(r)} - \frac{r_{max} - r_{min}}{\Delta r} \\ \epsilon_b &= e^\beta \sum_{r=r_{min}}^{r_{max}} (e^{-2v} - 2e^{-v}) + \frac{r_{max} - r_{min}}{\Delta r} \end{aligned} \tag{10}$$

where $\Delta r \to 0$ is the bin size for r and $[r_{min}, r_{max}]$ is the numerical interval of integration. As a result,

$$\begin{aligned} \epsilon_u &= e^\beta p_u + R \\ \epsilon_b &= e^\beta p_b - R \end{aligned} \tag{11}$$

where

$$
\begin{aligned}
p_u &= \sum_{r=r_{min}}^{r_{max}} e^{-v(r)} \\
p_b &= \sum_{r=r_{min}}^{r_{max}} (e^{-2v} - 2e^{-v}) \\
R &= -\frac{r_{max} - r_{min}}{\Delta r}
\end{aligned}
\tag{12}
$$

Thus, according to Eq. 4, Eq. 5 and Eq. 8, the logarithm of the partition function for directed graphs can be written as

$$
\begin{aligned}
\log z_\nu = n_\nu (l_\nu \log \frac{\epsilon_0}{l_\nu} + d_{u,\nu} \log \frac{\epsilon_u}{\epsilon_0} + d_{b,\nu} \log \frac{\epsilon_b}{\epsilon_0} \\
- \log n_\nu) + (N - l_\nu n_\nu) \log \frac{r q_0}{N - l_\nu n_\nu}
\end{aligned}
\tag{13}
$$

The corresponding energy is given according to Eq. 6 by

$$
U_\nu = \frac{n_\nu d_{u,\nu} P_u e^\beta}{P_u e^\beta + R} + \frac{n_\nu d_{b,\nu} P_b e^\beta}{P_b e^\beta - R} \tag{14}
$$

From Eq. 7, with $V_\nu = n_\nu l_\nu$ and the total number of unidirectional edges $\Gamma_{u,\nu} = n_\nu d_{u,\nu}$ and bidirectional edges $\Gamma_{b,\nu} = n_\nu d_{b,\nu}$ appearing in the motif, the entropy of a directed graph is given by:

$$
\begin{aligned}
S_\nu = N \log \frac{r q_0}{N - V_\nu} + V_\nu \log \left[\frac{\epsilon_0}{l_\nu} \frac{N - V_\nu}{r q_0} \right] - n_\nu \log n_\nu \\
+ \Gamma_{u,\nu} \log \left[\frac{\epsilon_u}{\epsilon_0} \exp \left[-\beta (1 - \frac{R}{\epsilon_u}) \right] \right] \\
+ \Gamma_{b,\nu} \log \left[\frac{\epsilon_b}{\epsilon_0} \exp \left[-\beta (1 + \frac{R}{\epsilon_b}) \right] \right]
\end{aligned}
\tag{15}
$$

There are five terms appearing in this expression for the directed network entropy, and have the following properties

- The first three terms are independent of temperature. The first of these terms is proportional to the number of nodes in the network and increases as the number of nodes in the motif of type ν increases. The second term is proportional to the number of nodes contained within the clique of type ν and decreases as the size l_ν of the motif increases. The third term decreases with the increasing frequency of the motif of type ν and controls the distribution of motif frequencies.

- The final two terms give the separate entropy contributions from unidirectional and bidirectional edges. When the graph is strongly directed, i.e. there are no bidirectional edges, then the final term vanishes. When, on the other hand, the graph is weakly directed, i.e. there are few unidirectional edges, then the entropy approaches for the undirected graph.

3.5. Temperature

With thermodynamic entropy and energy obtained through the motif expansion, we explore how the motif structure determines the thermodynamic temperature, i.e. the rate of change of energy with entropy. When the motif composition varies with time, then the effects of these change can be monitored by computing and observing changes in temperature. Suppose that the graph-sequence $G_T = \{G_1, ...G_i, ...G_n\}$ represent a time-varying network system sampled at different times t_1,...,t_n. Let G_i be the network at time epoch t_i. We perform our motif-based analysis on such sequences of time-varying or dynamic graphs. Here, the graphs G_i are time-varying samples in which the node-set is fixed and the edge-set is time-varying. Examples of the graphs which satisfy these requirements and are studied later in this paper include financial market networks and gene regulatory networks. In both cases the node-set is fixed and the edge-set varies with time. The network structure fluctuations are associated with a measured temperature T_c defined as.

$$\frac{1}{T_c\left(G_i, G_j\right)} = \frac{dS}{dE} = \frac{S_{\nu_i} - S_{\nu_j}}{U_{\nu_i} - U_{\nu_j}} \tag{16}$$

We treat $\beta = 1/T$ as a hyper-parameter associated with the gas-model, and this corresponds to the ambient temperature. To make this explicit, we define the ambient temperature to be $T_0 = 1/\beta$. The overall temperature of the network is the sim of the ambient and fluctuation temperature, i.e.

$$T = T_0 + T_c = \frac{1}{k\beta} + \frac{U_{\nu_i} - U_{\nu_j}}{S_{\nu_i} - S_{\nu_j}} \tag{17}$$

To further develop the expression for the overall temperature, we compute

the change in entropy

$$\begin{aligned} S_{\nu_i} - S_{\nu_j} = & N \log \frac{N - V_{\nu_j}}{N - V_{\nu_i}} + \log \frac{\epsilon_0}{r q_0} (V_{\nu_i} - V_{\nu_j}) \\ & + V_{\nu_i} \log \left[\frac{N - V_{\nu_i}}{l_{\nu_i}} \right] - V_{\nu_j} \log \left[\frac{N - V_{\nu_j}}{l_{\nu_j}} \right] \\ & - n_{\nu_i} \log n_{\nu_i} + n_{\nu_j} \log n_{\nu_j} \\ & + \{ \Gamma_{u,\nu_i} - \Gamma_{u,\nu_j} \} \log \left[\frac{\epsilon_u}{\epsilon_0} \exp \left[-\beta (1 - \frac{R}{\epsilon_u}) \right] \right] \\ & + \{ \Gamma_{b,\nu_i} - \Gamma_{b,\nu_j} \} \log \left[\frac{\epsilon_b}{\epsilon_0} \exp \left[-\beta (1 + \frac{R}{\epsilon_b}) \right] \right] \end{aligned} \tag{18}$$

The change in energy can be similarly computed as

$$\begin{aligned} U_{\nu_i} - U_{\nu_j} = & \frac{P_u e^{\beta}}{P_u e^{\beta} + R} (n_{\nu_i} d_{u,\nu_i} - n_{\nu_j} d_{u,\nu_j}) \\ & + \frac{P_b e^{\beta}}{P_b e^{\beta} - R} (n_{\nu_i} d_{b,\nu_i} - n_{\nu_j} d_{b,\nu_j}) \end{aligned} \tag{19}$$

Thus the overall temperature can be computed from the changes in the motif edge composition from epoch to epoch as the network evolves with time.

4. Experiments

In this section, we use the proposed thermodynamic framework to analyze the time evolution of real-world complex networks and investigate the utility of the graph motif thermodynamic variables, i.e. entropy, energy, and temperature. First, we investigate the correlation between motif entropy and motif energy. Second, we explore the relationship between the motif temperature and the structure of the evolving network. Third, we use these thermodynamic quantities to analyse realistic time-evolving networks to explore whether they can effectively reveal the evolution of detailed network structure. Finally, we use the proposed thermodynamic framework to a) analyse financial network evolution and detect temporal anomalies associated with global financial or political crises, and b) to study developmental state detection in gene regulatory networks varying with time.

4.1. DataSet

Drosophila melanogaster Gene Networks: The Drosophila melanogaster Gene Networks [44] are extracted from a time series of gene expression data measured during full life cycle of drosophila melanogaster. The data follows the dynamics of 588 development genes along 66-time points spanning four different developmental stages. i.e. embryonic(1-30), larval(31-40), pupal(41-58), adult(59-66), which constitute a time-varying network.

NYSE Financial Networks: New York Stock Exchange (NYSE) database [45] is composed of 347 stock and their associated daily closing prices over 6004 transaction days from January 1986 to February 2011. To extract market networks, we closely follow [46]. We use a sliding time window of 28 days to obtain a moving closing price sequence for each stock. We regard each individual stock as a node of the time evolving network. At each time epoch, to determine the set of edges between nodes we compute the time-lagged correlation coefficient between the moving stock closing price sequences for each pair of stock. We search for the time-lag that gives the maximum correlation coefficient, and if this is in the top 5 percentile of the cumulative distribution we create a connecting edge. The direction of the edge is determined by the sign of optimal time lag. Using this procedure we construct 5976 financial network samples corresponding to different trading days.

Financial Networks for the Past Ten Years: We construct financial market networks covering the past ten years. These encompass 416 stock with their daily closing prices for 2639 transaction days from January 2010 to June 2020. Here we use a time window of 20 days (different from the 28-day time window in the NYSE Financial networks) and slide the window with time to obtain a network sequence as above.

4.2. Experimental Settings

We detect the motifs for each network as shown in Fig. 1. We then compute the motif entropy, energy and temperature for the different datasets. The parameter settings used are listed in Table 2. The physical parameters of the

Table 2: Parameters for different datasets

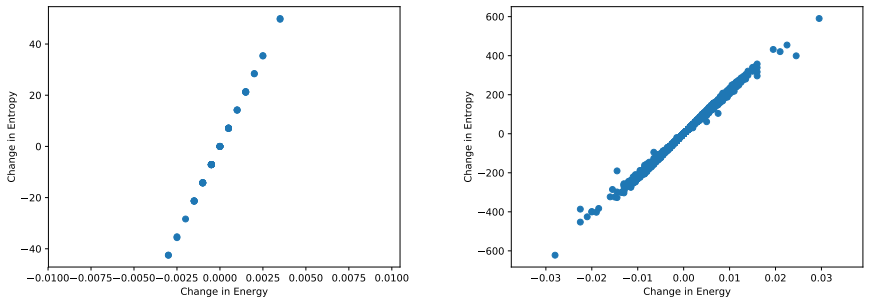
Datasets	Number of graph	Motifs used	Number of nodes	β	σ
Drosophila Networks	66	15	588	1000	3
NYSE Financial Networks	5976	15	347	100	9
Financial Networks for the Past Ten Years	2619	15	416	100	9

model are the inverse initial temperature β and the scale parameter σ appearing in the potential.

4.3. Correlation Between Entropy And Energy

In this subsection, we aim to explore the internal relationship between energy and entropy as they evolve with time. We first investigate these thermodynamic variables for the Drosophila melanogaster Gene Networks and the Financial Networks data. At each time step, we compute the motif entropy and motif energy for each graph. We construct two thermodynamic variable vectors, the motif entropy vector $S = (S_1, ..., S_t, S_{t+1}, ...)^T$ and the motif energy vector $U = (U_1, ...U_t, U_{t+1}, ...)^T$ for the entire network sequence. We then calculate the difference of motif entropy at the epochs t and $(t+1)$, which can be represented as $\delta S_t = S_t - S_{t+1}$, and the corresponding difference of motif energy is $\delta U = U_t - U_{t+1}$. We then construct a scatter plot of δS_t versus with δE for two realistic evolving networks for the samples spanned by the different time epochs.

The scatter plot for the Drosophila data is shown in Fig. 2 (a). The are a number of features to be noted from this data. Firstly, the number of distinct scatter points displayed in the figure is less than the number of Drosophila network samples over time. The reason for this is that many networks have the same change in motif energy and entropy compared to their previous sampling epoch. Secondly, the scatter points are arranged in an approximately straight line. When the change in energy equals zero, the change in entropy is also zero. According to the Eq. 16 the ratio of the change in entropy to the change in energy is the inverse fluctuation temperature. This result shows that when the Drosophila networks evolve with time, the motif entropy and energy change in



(a) Drosophila melanogaster

(b) Financial Networks

Figure 2: Figures (a) and (b) show the change in motif entropy versus the change in motif energy on Drosophila melanogaster networks and Financial Networks respectively.

a manner consistent with small fluctuations in the internal temperature. In Fig. 2 (b), we repeat this analysis for the financial market data. Again, most of the points lie on a tight regression line, of approximately constant slope. This indicates that the ratio of the change in entropy to the change in energy, i.e. the temperature, appears to show small fluctuations over time. There are also some points that deviate significantly from the regression trend line, which correspond to abrupt changes in the temperature associated with the network. These are associated with temporal anomalies in network structure, and are caused by external events such as global financial and political crises.

4.4. Temperature and Network Structure

In this section, we explore the relationship between the thermodynamic temperature variables and the change of network structure. To this end, we commence by constructing multiple groups of networks with different node degrees. We construct these graphs in a similar way to that in [47]. First, we construct a complete graph with 80 nodes, we continue to delete its edges randomly with a probability $p \in [0, 02]$. Second, we begin with the same complete graph, this time we delete edges randomly with the probability of $p + \triangle p$. Using these two random graphs, we compute the temperature according to Eq. 17. We repeat

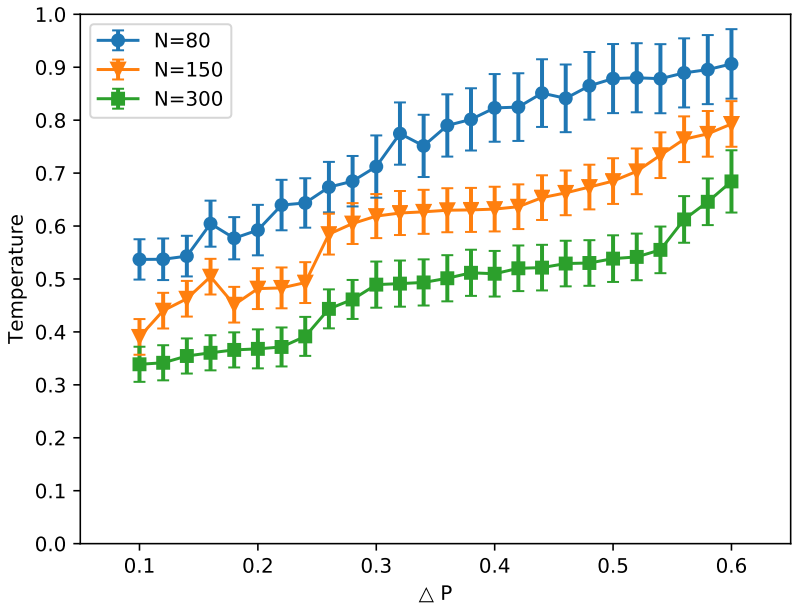


Figure 3: Mean and standard deviation of the temperature versus $\triangle p$ for random graphs with different graph sizes.

the process above 100 times with the variable probability of deleting edges $\triangle p$ varying from 0.1 to 0.6. In this way we construct a group of random networks consisting of 80 nodes, and the varying network structure which evolves with the change of node degree. We also construct the random graphs with 150 nodes and 300 nodes respectively. We plot the mean with the standard deviation(shown an error bar) of the temperature versus $\triangle p$ for the set of random graphs of different sizes.

As shown in Fig. 3, as $\triangle p$ increases, both the mean values and the variance values of temperature increases for all three kinds of graphs. This is because when there is a dramatic structural change in the time-varying network, the variance of the ratio $(S_{\nu_i} - S_{\nu_j})/(U_{\nu_i} - U_{\nu_j})$ becomes larger, therefore the variance of the temperature changes larger.

To take our study one step further, we investigate the relationship between

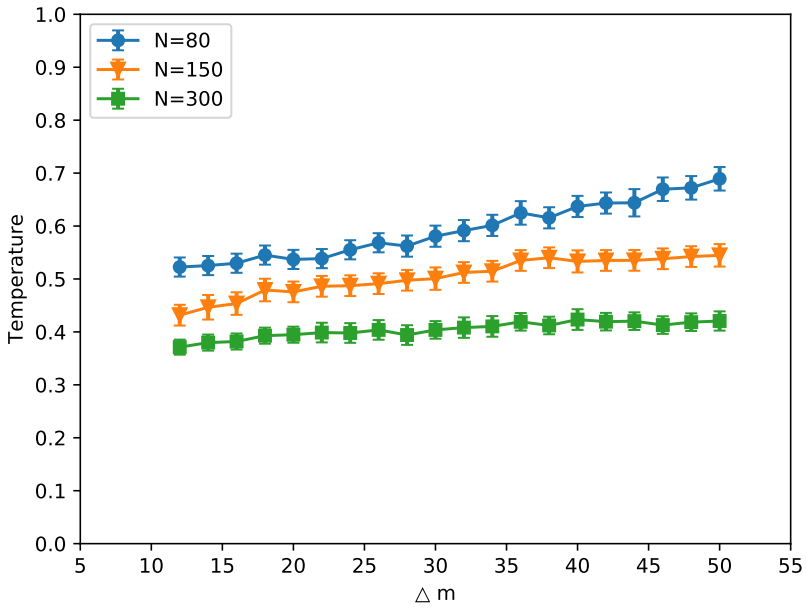


Figure 4: Mean and standard deviation of the temperature versus $\triangle m$ for regular graphs with different graph sizes.

temperature and network structure especially when the internal structure of the network has changed by a small amount during the evolution. We first construct a regular graph of 80 nodes with degree $m = 10$, and create a second regular graph with degree $m + \triangle m$. Thus the temperature difference between these two networks can be computed. We also create 100 realizations of these graphs with $\triangle m$ varying from 12 to 50. We plot the mean and standard deviation of the temperature versus $\triangle m$ for different sizes of graphs in Fig. 4. For each set of $\triangle m$, the variance value of the temperature stays almost constant. The reason for this is that each set of regular graphs with fixed node degree $m+\triangle m$, the network structure remains unchanged. The mean value of the temperature shows some small fluctuations with varying $\triangle m$. This is because the internal motif structure of the network changes in a regular way with the addition of two new neighbors to each node for each time.

4.5. Thermodynamic Measures for Analysing Network Evolution

To evaluate our thermodynamic framework in characterizing networks, we compute the thermodynamic motif entropy and motif energy respectively. We also compute the relative temperature between networks at consecutive time steps. We investigating how these thermodynamic network variables evolve with time, and evaluate their effectiveness in characterizing distinct periods in network evolution.

We compare the thermodynamic quantities computed using our motif method with the von Neumann entropy and associated thermodynamic variables described in [48]. Here the von Neumann entropy for a directed graph is given by

$$S_{VN}^{D} = 1 - \frac{1}{|\mathcal{V}|} - \frac{1}{2|\mathcal{V}|^2}\left\{\sum_{(u,v)\in\mathcal{E}_1}\frac{d_u^{in}}{d_v^{in}d_u^{out^2}} + \sum_{(u,v)\in\mathcal{E}_2}\frac{1}{d_u^{\text{out}}d_v^{\text{out}}}\right\},$$

where $\mathcal{V}$ is the number of nodes in the graph. The symbols $\mathcal{E}_1$ and $\mathcal{E}_2$ denote two disjoint subsets of the edge set of the graph, containing unidirectional and bidirectional edges respectively, and so $\mathcal{E}_1 = \{(u,v) \mid (u,v) \in \mathcal{E} \wedge (v,u) \notin \mathcal{E}\}$, $\mathcal{E}_2 = \{(u,v) \mid (u,v) \in \mathcal{E} \wedge (v,u) \in \mathcal{E}\}$ satisfy the conditions $\mathcal{E}_1 \cup \mathcal{E}_2 = \mathcal{E}, \mathcal{E}_1 \cap \mathcal{E}_2 = \emptyset$. For the internal energy $U = \sum_{s=1}^{|\mathcal{V}|} p_s E_s$ where E_s is the energy of microstate s and p_s is the probability that the system occupies a microstate indexed by s. The internal energy is taken to be the total number of edges in the graph i.e., $U = |\mathcal{E}|$. Finally the von Neumann temperature is obtained through

$$\frac{1}{T\left(\mathcal{G}_1, \mathcal{G}_2\right)} = \sum_{(u,v)\in\mathcal{E}_1,\mathcal{E}_2}\frac{d_u\Delta_v + d_v\Delta_u + \Delta_u\Delta_v}{\Delta|\mathcal{E}|d_u\left(d_u+\Delta_u\right)d_v\left(d_v+\Delta_v\right)},$$

where$\Delta|\mathcal{E}| = |\mathcal{E}_2| - |\mathcal{E}_1|$, Δ_u and Δ_v are defined as the difference between the degree of vertex u in graphs G_2 and G_1.

In Figure 5 and Figure 6, we compare our thermodynamic framework with the von Neumann entropy and the average node degree, which is a widely used measure of network structure. The four developmental stages for Drosophila are shown in different colors (name them here embryonic, larval, pupal, and adult).

From Fig. 5, it is clear that von Neumann entropy shows three different evolutionary periods where it is either increasing, decreasing, or constant.

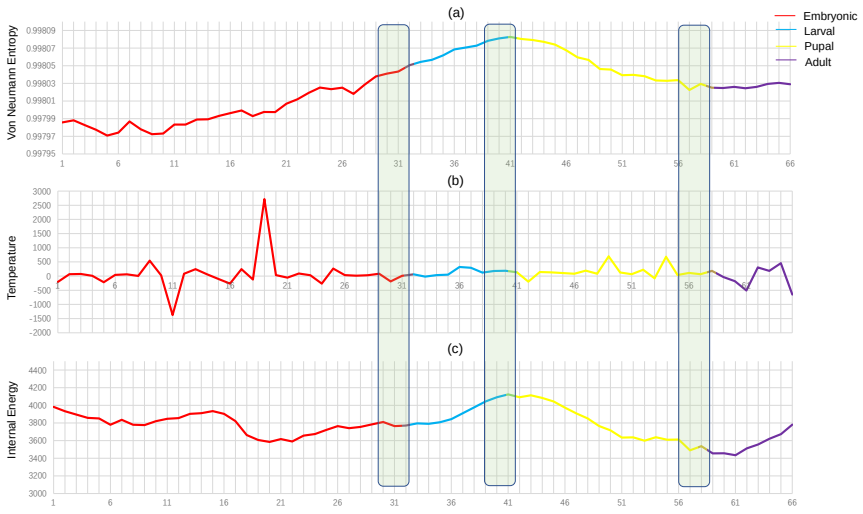


Figure 5: Top to bottom: the temporal evolution of (a) the von Neumann entropy, (b) the temperature, and (c) the internal energy for the dynamic Drosophila melanogaster network.

These three states successfully distinguish the pupal and adult periods in the Drosophila development, but do not distinguish between the embryonic and larval periods. For the motif entropy, on the other hand, in Fig. 6, the section marked in red (embryonic period) shows some local fluctuations due to structural changes associated with the early development of the embryo. However, the subsequent three stages show obvious increasing and decreasing trends, with which we can clearly distinguish the different developmental periods for Drosophila.

The temperature is computed as the partial derivative of energy and with respect to entropy. Here we focus on the boundaries between different developmental periods. Comparing the temperature in Figs. 5 and Fig. 6, we can find that when the growth period of Drosophila changes, the motif temperature curve shows obvious transitions, while the von Neumann temperature remains almost unchanged. This result indicates that the motif temperature can well represent the transitions between different developmental periods for the Drosophila from the gene regulatory network data.

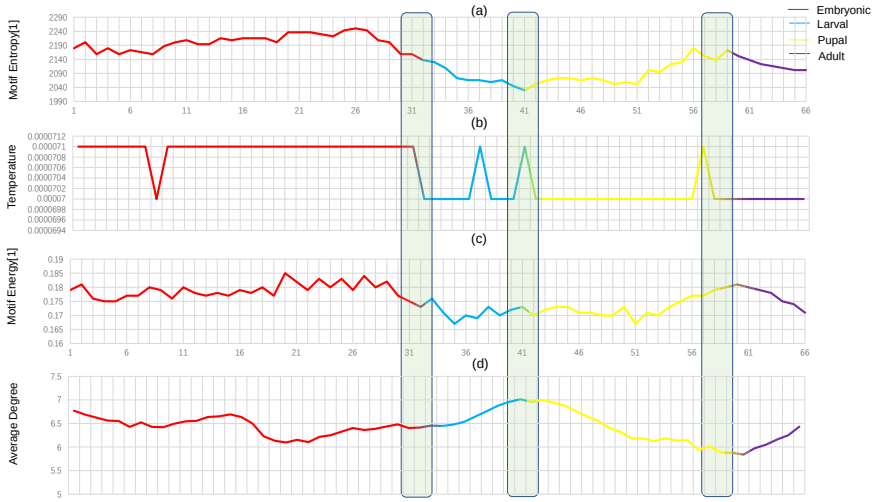


Figure 6: Top to bottom: the temporal evolution of (a) the motif entropy, (b) the temperature, (c) the motif energy, and (d) the average degree for the dynamic Drosophila melanogaster network.

Finally, the internal energy shown in Fig. 5 is equal to the number of edges in each graph, which is proportional to the average degree shown in Fig. 6. The reason for this is that the average degree is computed from the number of edges in the graph, and the number of nodes in the graph remains unchanged over time. Both the internal energy and the motif energy behave in a similar manner to the corresponding entropy.

4.6. Thermodynamic Framework for Characterizing Abrupt Changes in Financial Networks

4.6.1. Financial Networks for the Past Ten Years

We continue our study by exploring whether the motif thermodynamic framework can be used for a better understanding of the time evolution of financial networks. To this end, we first compare the evolutionary behavior of motif entropy and von Neumann entropy for the financial networks. At each time step, we compute both the motif entropy and von Neumann entropy. For a graph G with adjacency matrix A, according to Passerini and Severini [49] the

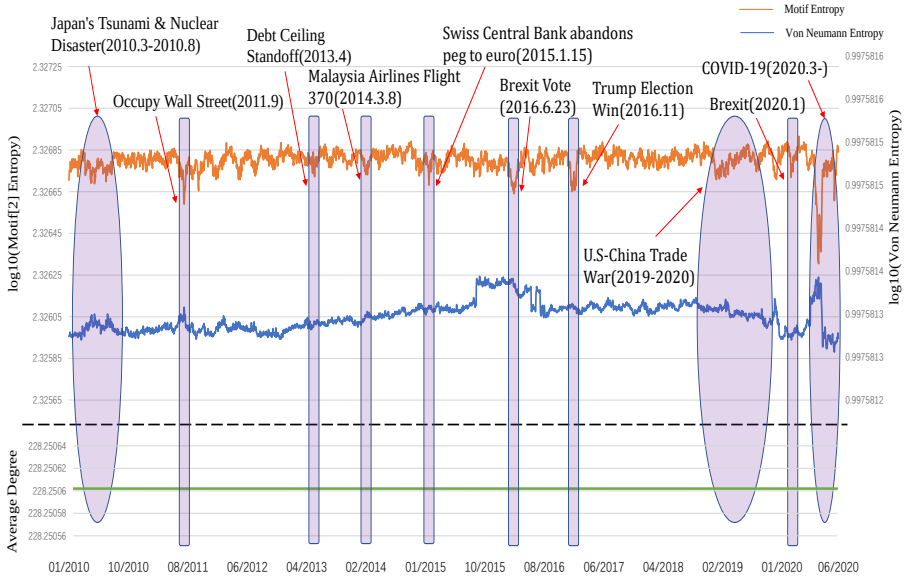


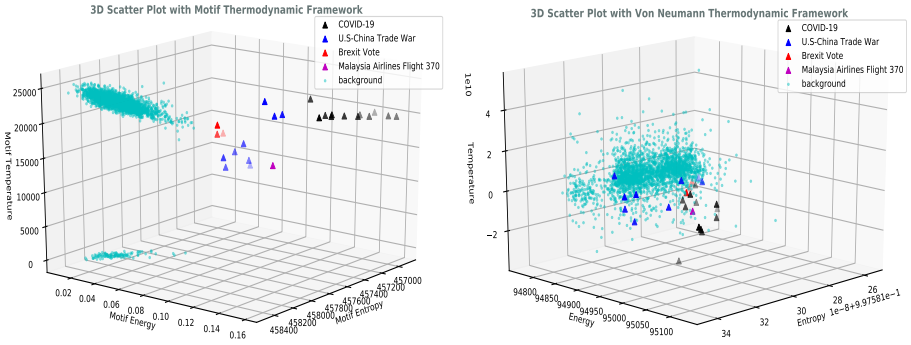
Figure 7: The motif entropy (upper, orange), von Neumann entropy (middle, blue), and average degree (lower, green) versus time for the financial networks covering past ten years.

von-Neumann entropy is

$$S_{vN}(G) = -Tr[\frac{\tilde{L}}{|V|}\ln\frac{\tilde{L}}{|V|}]$$

where $\tilde{L} = D^{-1/2}(D-A)D^{-1/2}$ is the normalised Laplacian matrix of the graph.

We calculate the second motif entropy and the von Neumann entropy for each sample of the graph and plot the two entropy values for the stock network as it evolves with time. We choose the value of second motif entropy instead of the first motif entropy since the networks are mainly composed of the second motif. Figure 7 shows the time series for the motif entropy (orange line), the von Neumann entropy (blue line), and the average degree (green line) for 2619 trading day. When a financial crisis occurs, there is a sudden drop (i.e. a trough) in the motif entropy. Examples include the Debt Ceiling Standoff (2013.4), the Brexit Vote (2016.6.23), the Trump Election Win (2016.11). In the case of the von Neumann entropy, it is hard to distinguish these events on the basis of entropy. On the other hand, for some crisis events which persist



(a) (b)

Figure 8: The 3D scatter plot of the financial networks in the thermodynamic space. The financial crisis period is represented by triangular symbols of different colors, while the remaining periods are represented by green dots, and these form the background. (a) the thermodynamic framework spanned by motif entropy, motif energy and motif temperature. (b)the thermodynamic framework spanned by von Neumann entropy, internal energy and temperature.

over an extended period, e.g. the U.S-China Trade War (2019-2020), the motif entropy reveals multiple anomalies over the period. The von Neumann entropy shows fluctuations that are similar to those exhibited over crisis free periods. In before, we find that the average degree of the graph remains constant throughout time so that we cannot capture any useful information through this classical graph structure measurements. This is because when we construct the graph, we choose to establish edge connections between the nodes whose correlation coefficient is in the top 5 percentile. Since the number of nodes in the graph does not change with the network, the number of edges remains the same, and the average degree of the network eventually stays constant.

To further explore whether the thermodynamic motif framework can be used as an effective tool for better understanding the evolution of financial networks, we calculate the motif entropy, motif energy, and the corresponding temperature for each graph evolving with time, and plot them as a three-dimensional scatter plot in Fig. 8(a). We also create a three-dimensional scatter plot for the

von Neumann thermodynamic framework with von Neumann entropy, internal energy, and corresponding temperature as the coordinates, the result is shown in Figure 8(b).

As is shown in Fig. 8,when we use the motif-based framework the entire time series during the 2619 trading days and multiple financial crises shown with different colored markers are clearly separated. On the other hand, in the case of the von Neumann thermodynamic variables the graphs corresponding to different time epochs are badly interspersed. On the other hand, for the ambient (non-crisis) time represented by green dots, the motif thermodynamic framework also performs better than the von Neumann thermodynamic framework because it gives rise to a strong manifold structure with good temporal continuity.

4.6.2. NYSE Financial Networks Anomalous Event Detection

To see more clearly the detail of how the thermodynamic variables change over time during the different periods and test the ability of the thermodynamic methods to characterize a network when it encounters network a sudden change in structure, we explore how the motif entropy of the NYSE stock networks evolve with time. We select nine different financial crises(i.e. Black Monday, Friday the 13th Mini Crash, Early 1990s recession, Asian Financial Crisis, Russian Financial Crisis, Dot-com Bubble, September 11 attacks, Downturn of 2002-2003, Financial Crisis of 2007-2008). We compare our motif entropy with the von Neumann entropy. As we can see in Fig. 9,the motif entropy responds strongly to the crises while varying smoothly before and after the crisis period. The von Neumann entropy, on the other hand, fails to detect several financial crises. In the case of the Early 1990s Recession events though, the von Neumann entropy has a strong shoulder.

4.6.3. Time Series Embeddings

To take our study one step further we focus on detecting temporal anomalies. To do this we perform embeddings of the financial networks time series covering past ten years. For the entropy vectors of the networks, we use kernel

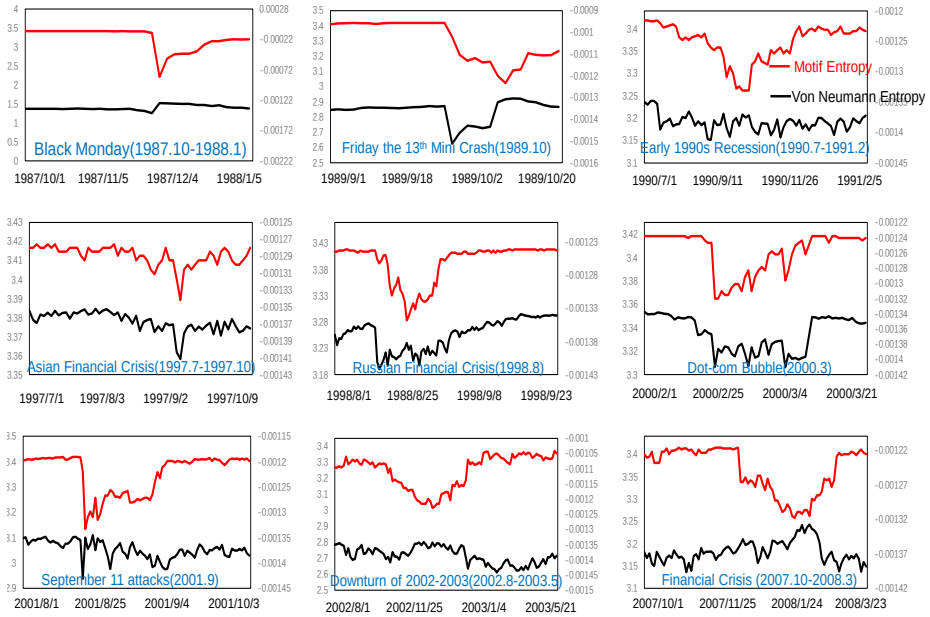


Figure 9: The evaluation of motif entropy (upper red) and von Neumann entropy(lower black) for nine different global events in Financial networks

PCA. We use the graph entropies to construct a kernel matrix, and then perform linear kernel principal components analysis to embed the sample of graphs into a vector space. Let H be the matrix of entropy differences with element $H(i,j) = ||S_{VN}(i) - S_{VN}(j)||$. We use the entropy similarity matrix to compute a symmetric kernel matrix

$$K = -1/2(I - J/M)H(I - J/M)$$

where I is the $M \times M$ identity matrix and $J = ee^T$ where $e = (1, 1, ..., 1)^T$ is the all-ones vector of length M. We perform kernel embedding on the matrix K. To this end let Y be the matrix with the embedding coordinates of the graphs as columns, then $K = X^T X$. Performing the eigendecomposition $K = U\Lambda U^T$, the matrix of embedding coordinates is $X = \sqrt{\Lambda}U^T$. We visualize the distribution of the graphs using the first three rows of X corresponding to the leading three eigenvectors of the kernel matrix.

Figure 10 illustrates the structural changes in the embedding spaces before

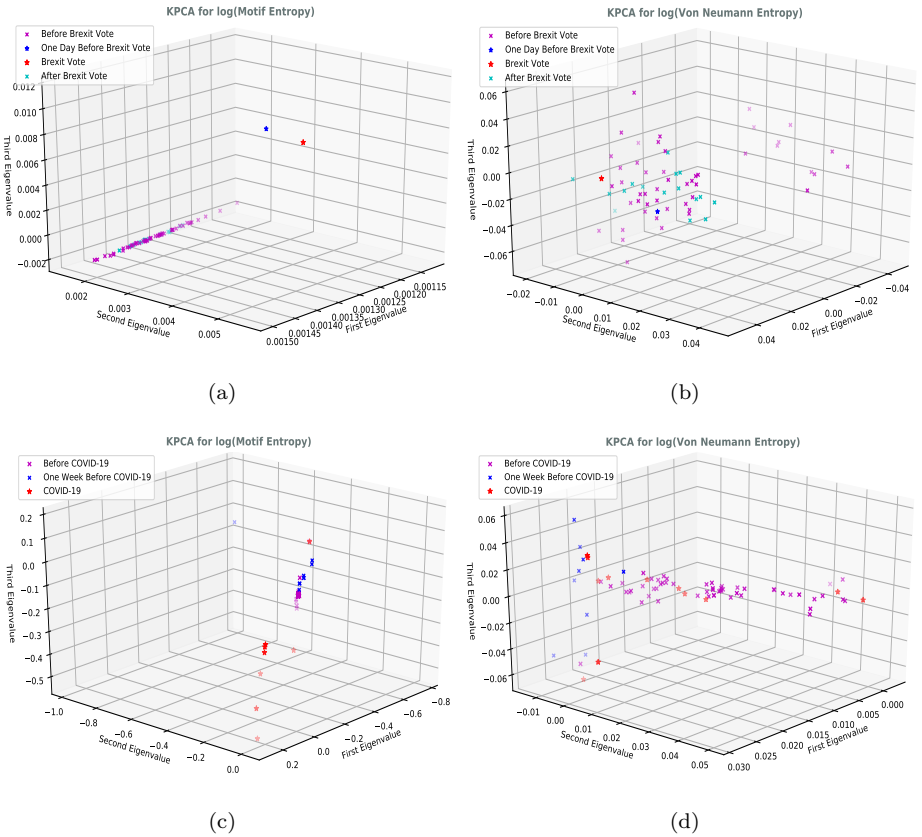


Figure 10: Embeddings in the proximity of the Black Monday event. Figure 10 (a) and (b) depict the spatial distribution of embeddings obtained by KPCA of the Motif entropy vectors and the von Neumann entropy during the UK Brexit Vote, respectively. While Figure 10 (c) and (d) illustrates the distribution around the period of the COVID-19.

and after crucial events generated from motif entropy and Von Neumann entropy, respectively. The blue star represents the exact day before the crisis occurred while the red stars represent the period during the crises. From the figure, it is clear that both types of event can be effectively detected by the two entropies. From Figure (a)(b), it is clear that the crisis event can be effectively detected by motif entropy while the von Neumann entropy failed, and the motif entropy keeps the continuity in time. From Figure (c) and (d), we can find that the motif entropy shows the changes before the event happens(blue points)

and multiple exceptions during event occurrence(red stars). This is because the COVID-19 is a long-term global crisis event that persists until the end of the data.

5. Conclusion

In this paper, we proposed a novel thermodynamic framework for the analysis of time-varying networks based on motif. Commencing from constructing the relationship between network structure and gas system structure, we introduce the partition function for the thermodynamic system. We continue to compare the cluster in the gas system to the motifs in directed networks and derive the partition function based on the motif for the network. Then we compute the motif entropy and motif energy for the network through the motif partition function, the temperature can also be get with the partial derivative relation between energy and entropy. We finally applied our method on the financial networks, the result all demonstrate the effectiveness of three thermodynamic variables.

Our future work will focus on applying thermodynamic analysis to capture the internal laws of network evolution and predict network development. For example, we can predict the possible financial crisis events in the stock market, or predict whether the protein will have pathological changes. Another interesting line of investigation would be to explore if the thermodynamic framework can be extended to the domains of more complex networks for example the dynamic weighted graphs and study the characteristics of these thermodynamic variables.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China under Grant U2066213 and 61860206004, and in part by the Research Funds of State Grid Shaanxi Electric Power Company and State Grid Shaanxi Information and Telecommunication Company (contract no.SGSNXT00GCJS2000104).

Dongdong Chen and Xingchen Guo are co-first authors, which contributed equally to this paper.

References

[1] Z. Zhang, D. Chen, L. Bai, J. Wang, E. R. Hancock, Graph motif entropy for understanding time-evolving networks, IEEE Transactions on Neural Networks and Learning Systems (2020).

[2] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics 74 (1) (2002) 47.

[3] K. Anand, G. Bianconi, Entropy measures for networks: Toward an information theory of complex topologies, Physical Review E Statistical Nonlinear and Soft Matter Physics 80 (4 Pt 2) (2009) 045102.

[4] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (5594) (2002) 824–827.

[5] I. A.Maraziotis, S. Perantonis, A. Dragomir, D. Thanos, K-nets: Clustering through nearest neighbors networks, Pattern Recognition 88 (2019) 470–481.

[6] N. Rosenfeld, M. B. Elowitz, U. Alon, Negative autoregulation speeds the response times of transcription networks, Journal of Molecular Biology 323 (5) (2002) 785–793.

[7] A. Becskei, L. Serrano, Engineering stability in gene networks by autoregulation, Nature 405 (6786) (2000) 590–593.

[8] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, U. Alon, Superfamilies of designed and evolved networks, Science 303 (5663) (2004) 1538–1542.

[9] H. Bolouri, E. H. Davidson, Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics, Proceedings of the National Academy of Sciences 100 (16) (2003) 9371–9376.

[10] S. Hooshangi, S. Thiberge, R. Weiss, Ultrasensitivity and noise propagation in a synthetic transcriptional cascade, Proceedings of the National Academy of Sciences 102 (10) (2005) 3581–3586.

[11] Y. T. Maeda, M. Sano, Regulatory dynamics of synthetic gene networks with positive feedback, Journal of Molecular Biology 359 (4) (2006) 1107–1124.

[12] S. Kalir, S. Mangan, U. Alon, A coherent feed-forward loop with a sum input function prolongs flagella expression in escherichia coli, Molecular Systems Biology 1 (1) (2005) 2005–0006.

[13] S. Mangan, A. Zaslaver, U. Alon, The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks, Journal of Molecular Biology 334 (2) (2003) 197–204.

[14] M. Ronen, R. Rosenberg, B. I. Shraiman, U. Alon, Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics, Proceedings of The National Academy of Sciences 99 (16) (2002) 10555–10560.

[15] N. Rosenfeld, U. Alon, Response delays and the structure of transcription networks, Journal of Molecular Biology 329 (4) (2003) 645–654.

[16] I. Amit, A. Citri, T. Shay, Y. Lu, M. Katz, F. Zhang, G. Tarcic, D. Siwak, J. Lahad, J. Jacob-Hirsch, et al., A module of negative feedback regulators defines growth factor signaling, Nature Genetics 39 (4) (2007) 503–512.

[17] E. Wong, B. Baur, S. Quader, C.-H. Huang, Biological network motif detection: principles and practice, Briefings In Bioinformatics 13 (2) (2011) 202–215.

[18] U. Alon, Network motifs: theory and experimental approaches, Nature Reviews Genetics 8 (6) (2007) 450–461.

[19] J. Wang, R. C. Wilson, E. R. Hancock, Spin statistics, partition functions and network entropy, Journal of Complex Networks 5 (6) (2017) 858–883.

[20] C. Ye, R. Wilson, L. Rossi, A. Torsello, E. Hancock, Thermodynamic analysis of time evolving networks, Entropy 20 (10) (2018) 759.

[21] Q. Hui, Y. Jihao, L. Xiaoyan, Lg: A clustering framework supported by point proximity relations, Pattern Recognition 103 (2020) 107265.

[22] J. C. Delvenne, A.-S. Libert, Centrality measures and thermodynamic formalism for complex networks, Physical Review E Statistical Nonlinear and Soft Matter Physics 83 (4) (2011) 046117.

[23] Y. Chen, T. T. Georgiou, M. Pavon, Ruelle-bowen continuous-time random walk, arXiv: Optimization and Control (2018).

[24] E. Estrada, N. Hatano, Statistical-mechanical approach to subgraph centrality in complex networks, Chemical Physics Letters 439 (1-3) (2009) 247–251.

[25] S. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of escherichia coli, Nature Genetics 31 (1) (2002) 64–68.

[26] Y. Sui, G. Wang, L. Zhang, Sparse subspace clustering via low-rank structure propagation, Pattern Recognition 95 (2019) 261–271.

[27] A. Awan, H. Bari, F. Yan, S. Moksong, S. Yang, S. Chowdhury, Q. Cui, Z. Yu, E. Purisima, E. Wang, Regulatory network motifs and hotspots of cancer genes in a mammalian cellular signalling network, IET Systems Biology 1 (5) (2007) 292–297.

[28] L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, D. B. Chklovskii, Structural properties of the caenorhabditis elegans neuronal network, PLoS Computational Biology 7 (2) (2011) e1001066.

[29] E. Yeger-Lotem, S. Sattath, N. Kashtan, S. Itzkovitz, R. Milo, R. Y. Pinter, U. Alon, H. Margalit, Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction, Proceedings of the National Academy of Sciences 101 (16) (2004) 5934–5939.

[30] T. Milenković, I. Filippis, M. Lappe, N. Pržulj, Optimized null model for protein structure networks, PLoS One 4 (6) (2009) e5967.

[31] M. P. Stumpf, C. Wiuf, R. M. May, Subnets of scale-free networks are not scale-free: sampling properties of networks, Proceedings of the National Academy of Sciences 102 (12) (2005) 4221–4224.

[32] M. Bressan, F. Chierichetti, R. Kumar, S. Leucci, A. Panconesi, Motif counting beyond five nodes, Acm Transactions on Knowledge Discovery from Data 12 (4) (2018) 1–25.

[33] K. Baskerville, P. Grassberger, M. Paczuski, Graph animals, subgraph sampling and motif search in large networks, Physical Review E Statal Nonlinear And Soft Matter Physics 76 (3 Pt 2) (2007) 036107.

[34] E. Wijaya, K. Rajaraman, S. M. Yiu, , W.-K. Sung, Detection of generic spaced motifs using submotif pattern mining, Bioinformatics 23 (12) (2007) 1476–1485.

[35] H. Li, X. Liu, T. Li, R. Gan, A novel density-based clustering algorithm using nearest neighbor graph, Pattern Recognition 102 (2020) 107206.

[36] J. E. Mayer, The statistical mechanics of condensing systems. i, The Journal of Chemical Physics 5 (1) (1937) 67–73.

[37] B. Kahn, G. E. Uhlenbeck, On the theory of condensation, Physica 5 (5) (1938) 399–416.

[38] T. D. Lee, C. N. Yang, Many-body problem in quantum statistical mechanics. i. general formulation, Physical Review 113 (5) (1959) 1165.

[39] B. Halperin, D. R. Nelson, Theory of two-dimensional melting, Physical Review Letters 41 (2) (1978) 121.

[40] A. Young, Melting and the vector coulomb gas in two dimensions, Physical Review B 19 (4) (1979) 1855.

[41] N. D. Mermin, Erratum: Crystalline order in two dimensions, Physical Review B 20 (11) (1968) 4762–4762.

[42] L. Cao, C. Li, T. Mueller, The use of cluster expansions to predict the structures and properties of surfaces and nanostructured materials, Journal of Chemical Information and Modeling 58 (12) (2018) 2401–2413.

[43] J. L. Gout, A. Guessab, Methods of numerical integration, Esaim Mathematical Modelling And Numerical Analysis 20 (2) (1986) 287–314.

[44] M. Arbeitman, E. E. Furlong, F. Imam, E. Johnson, B. Null, B. Baker, M. Krasnow, M. Scott, R. Davis, K. White, Gene expression during the life cycle of, Science (New York, N.Y.) 297 (2002) 2270–5. doi:10.1126/science.1072152.

[45] F. N. Silva, C. H. Comin, T. K. D. Peron, F. A. Rodrigues, Y. Cheng, R. C. Wilson, E. Hancock, L. D. F. Costa, Modular dynamics of financial market networks, Quantitative Finance (2015).

[46] J. Wang, C. Lin, Y. Wang, Thermodynamic entropy in quantum statistics for stock market networks, Complexity 2019 (2019) 1–11. doi:10.1155/2019/1817248.

[47] C. Ye, C. H. Comin, T. K. D. M. Peron, F. N. Silva, F. A. Rodrigues, L. da Fontoura Costa, A. Torsello, E. R. Hancock, Thermodynamic characterization of networks using graph polynomials, Physical Review. E 92 3 (2015) 032810.

[48] C. Ye, R. C. Wilson, C. H. Comin, L. d. F. Costa, E. R. Hancock, Approximate von neumann entropy for directed graphs, Physical Review E 89 (5) (2014) 052804.

[49] F. Passerini, S. Severini, The von neumann entropy of networks, International Journal of Agent Technologies and Systems 1 (4) (2008) 58–67.

Dongdong Chen is now a postgraduate student at the School of Informatics Xiamen University, China. His research interests include data mining, machine learning, and network representation learning.

Xingchen Guo received his B. Sc. degree (2015) in Mechanical Design Manufacture and Automation from Shaanxi University of Science & Technology, China. He is currently a Master student in the School of Management at Xi'an Jiaotong University, Xian, China. His research interests include pattern recognition and machine learning, particularly problems involving graphs and networks.

Jianjia Wang received the B.Sc. degree from Nanjing University of Posts and Telecommunications (2011) and M.Sc. degrees from Hong Kong University of Science and Technology (2013). He worked as a research assistant at Hong Kong Applied Science and Technology Research Institute from 2013 to 2014. He is currently pursuing the Ph.D. degree in the Department of Computer Science, University of York, U.K. His research interests include statistical pattern recognition, complex networks, information theory, thermodynamic and quantum statistics, especially in graph and network analysis.

Jiatong Liu is now a postgraduate student at the School of Informatics Xiamen University, China. Her research interests include data mining, machine learning, and network representation learning.

Zhihong Zhang received his BSc degree (1st class Hons.) in computer science from the University of Ulster, UK, in 2009 and the PhD degree in computer science from the University of York, UK, in 2013. He won the K. M. Stott prize for best thesis from the University of York in 2013. He is now an associate professor at the School of Informatics Xiamen University, China. His research interests are wide-reaching but mainly involve the areas of pattern recognition and machine learning, particularly problems involving graphs and networks. He is a recipient of the Best Paper Awards of the International Conference on Pattern Recognition ICPR 2018. He is currently an Associate Editor of Pattern Recognition Journal.

Edwin R. Hancock holds a BSc degree in physics (1977), a PhD degree in high-energy physics (1981) and a D.Sc. degree (2008) from the University of Durham, and a doctorate Honoris Causa from the University of

Alicante in 2015. He is currently Emeritus Professor in the Department of Computer Science at the University of York, and also Adjunct Professor and Principal Investigator - Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University. He became a fellow of the International Association for Pattern Recognition in 2000, the Institute of Physics in 2007 and the IEEE in 2016. He is currently Editor-in-Chief of the journal Pattern Recognition, and was founding Editor-in-Chief of IET Computer Vision from 2006 until 2012. He has also been a member of the editorial boards of the journals IEEE Transactions on Pattern Analysis and Machine Intelligence, Pattern Recognition, Computer Vision and Image Understanding, Image and Vision Computing, and the International Journal of Complex Networks. He was Conference Chair in 1994 and Programme Chair in 2016 for the British Machine Vision Conference, Track Chair for ICPR (2004 and 2016) and Area Chair for ECCV (2006) and CVPR (2008 and 2014). In 1997 he jointly established the EMMCVPR workshop series with Marcello Pelillo. He was awarded a Royal Society Wolfson Research Merit Award in 2009, was named BMVA Distinguished Fellow in 2016 and received the IAPR Pierre Devijver Award in 2018. He was a Governing Board Member of the IAPR (2006-2016), and was Second Vice President of the Association (2016-2018). He was a member of the Computer Science and Information Technology sub-panel for the UK REF 2014 and will also be a panelist for REF 2021. He has published about 200 journal papers and 650 refereed conference papers.