

Heat Flow-Thermodynamic Depth Complexity in Directed Networks

Francisco Escolano¹, Boyan Bonev¹, and Edwin R. Hancock²

¹ University of Alicante
{sco,boyan}@dccia.ua.es
² University of York
erh@cs.york.ac.uk

Abstract. In this paper we extend the heat diffusion-thermodynamic depth approach for undirected networks/graphs to directed graphs. This extension is motivated by the need to measure the complexity of structural patterns encoded by directed graphs. It consists of: a) analyzing and characterizing heat diffusion traces in directed graphs, b) extending the thermodynamic depth framework to capture the second-order variability of the diffusion traces to measure the complexity of directed networks. In our experiments we characterize several directed networks derived from different natural languages. We show that our proposed extension finds differences between languages that are blind to the classical analysis of degree distributions.

1 Introduction

The quantification of the complexity of patterns plays a fundamental role in pattern recognition and machine learning. Information theory [1] provides principled approaches to the analysis complexity that include minimum description length (MDL) and minimum message length (MML) which allow us to find the model that parsimoniously describes vectorial data. However, the latter principles have not been incorporated to the graph domain until recently (see [2] for trees and [3] for edge-weighted undirected graphs). In fact, the intersection between structural pattern recognition and complex networks has proved to be fruitful and has inspired several interesting measures of graph complexity. Most of these measures rely on quantifying the degree of randomness of the structural representation. For instance, Körner entropy was motivated by the need to measure how much information can flow through a graph, when pairs of symbols can be confused [4]. This implies admitting a probability that a memoryless source emits a symbol. For each node in the graph there is a symbol, and two nodes are adjacent if their symbols are distinguishable. In this setting entropy is defined as the minimal cross entropy between the probability distribution and the vertex packing polytope of the graph. Since the vertex packing polytope is the convex hull of all characteristic vectors of stable sets of the graph, the task of measuring Körner entropy relies on solving an NP problem. More recently, Passerini and Severini have applied the quantum (von Neumann) entropy to graphs [5]. The state of a quantum mechanical system of a finite dimension is defined by a

density matrix for networks that can be modeled the combinatorial Laplacian. As a result the von Neumann entropy is given by the Shannon entropy of the Laplacian eigenvalues (normalized by the volume of the graph). This definition of graph entropy is maximal for random graphs, minimal for complete ones and intermediate for star graphs. Although the computation of the von Neumann entropy is cubic with the size of the network, it has recently been shown how this can be approximated using quadratic degree statistics and successfully applied to structural discrimination [6].

The Körner and von Neumann entropies are two examples of *randomness* complexity applied to graphs. An alternative is to use the so called *statistical* complexity and to quantify the regularities of the structure beyond its randomness [7]. The general underlying principle of statistical complexity is that it is zero for both random and regular/complete (completely ordered) graphs. A very recent example of statistical complexity is the Estrada heterogeneity index [8]. This index is defined as the Dirichlet sum of root squared degree differences. The obtained index is related to the Randić index [9]. Estrada's heterogeneity index is obviously zero for both random graphs and regular or complete ones. Another method which goes beyond randomness complexity is our recent application of thermodynamic depth [10] to the domain of graphs [11]. This involves defining both the macro-states (the graph) and the micro-states (the nodes). Complexity is quantified in terms of the amount of structural variability as a node evolves through a subgraph containing adjacent nodes (first order expansion) and eventually encompasses the full graph (if possible). Each expansion step is characterized by the temporal trace of heat flowing through the network. The sequence of expansions for a given node is referred to as a *history* [12], which contains the heat flow traces of each node. The variability of a given history quantifies how uniformly the full graph emerges from the corresponding node. The average heat flow traces of all the nodes can be combined to yield a second-order variability measure, the so called thermodynamic depth of the graph. Shallow (low-variability) graphs are characterized by similar histories with low variance and this means that heat flows satisfies similar topological constraints at each node. In contrast, deep graphs emerge from histories with large variance. Both random and complete graphs have zero depth, whereas grids and linear graphs have larger depths. A nice property of our thermodynamic depth approach is that it can be applied not only to heat flow traces but also to the heterogeneity index and the von Neumann entropy. It has been successfully used to correlate PPI networks with the phyla of bacteria.

All of the above approaches are confined to undirected graphs since many networks or graphs in the real world can be modeled with them: (e.g. protein-protein interaction (PPI) networks, shapes as Delaunay triangulations, adjacency graphs in images). However, considering the orientation of edges (e.g. directed trees and causal graphs in Bayesian networks) adds meaningful information which allows us to model networks such as metabolic pathways (cascades of chemical reactions) as well as natural languages (where the relative order of words matters) together with social networks (e.g. citation networks). The Internet is a clear

example of a directed network and Pagerank is an example of well known algorithm [13] which exemplifies the formal difficulty of analyzing directed graphs and the study of how the information flows through them in the context of the Internet. When a sink node (a node with zero outdegree) is reached by a random walk, there is a given (small) probability of making a transition to any other node in the network (this is called *teleporting*). The Laplacian of a directed graph can be defined through a symmetrization process provided that the transition matrix for the random walks allows for teleporting. Given a symmetric semi-definite operator such as the directed Laplacian, it is straightforward to compute heat kernels and thus to evaluate diffusion flow traces. A natural way of quantifying the complexity of directed graphs is to extend our thermodynamic depth approach to deal with oriented edges.

In this paper we will address that challenging point as follows. In Section 2 we will describe the directed Laplacian. Section 3 is devoted to the analysis of the fundamental formal differences between undirected and directed graphs in terms of heat flow diffusion using heat kernels. In Section 4 we redefine the thermodynamic depth for digraphs. Experiments and discussion (Section 5) are focused on the analysis of directed networks derived from natural languages and the quantification of their complexities. For instance, we show that our proposed extension can identify differences between languages that are blind to the classical analysis of degree distributions. We conclude this paper with a summary of our contributions and suggestions for future work.

2 The Laplacian of a Directed Graph

A directed graph (digraph) $G = (V, E)$ with $n = |V|$ vertices and edges $E \subseteq E \times E$ is encoded by an adjacency matrix \mathbf{A} where $A_{ij} > 0$ if $i \rightarrow j \in E$ and $A_{ij} = 0$ otherwise (this definition includes weighted adjacency matrices). The outdegree matrix \mathbf{D} is a diagonal matrix where $D_{ii} = \sum_{j \in V} A_{ij}$. The transition matrix \mathbf{P} is defined by $P_{ij} = \frac{A_{ij}}{D_{ii}}$ if $(i, j) \in E$ and $P_{ij} = 0$ otherwise. The transition matrix is key to defining random walks on the digraph and P_{ij} is the probability of reaching node j from node i . Given these definitions we have that $\sum_{j \in V} P_{ij} \neq 1$ in general. In addition, \mathbf{P} is irreducible iff G is strongly connected (there is path from each vertex to every other vertex). If \mathbf{P} is irreducible, the Perron-Frobenius theorem ensures that there exists a left eigenvector ϕ satisfying $\phi^T \mathbf{P} = \lambda \phi^T$ and $\phi(i) > 0 \forall i$. If \mathbf{P} is aperiodic (spectral radius $\rho = 1$) we have $\phi^T \mathbf{P} = \rho \phi^T$ and all the other eigenvalues have an absolute value smaller than $\rho = 1$. By ensuring strong connection and aperiodicity we also ensure that any random walk in a directed graph satisfying these two properties converges to a unique stationary distribution.

Normalizing ϕ so that $\sum_{i \in V} \phi(i) = 1$, we encode the eigenvector elements as a probability distribution. This normalized row vector ϕ corresponds to the stationary distribution of the random walks defined by \mathbf{P} since $\phi \mathbf{P} = \phi$. Therefore, $\phi(i) = \sum_{j, j \rightarrow i} \phi(j) P_{ji}$, that is, the probability of that the random walk is at node i is the sum of all incoming probabilities from all nodes j satisfying $j \rightarrow i$.

If we define $\Phi = \text{diag}(\phi(1) \dots \phi(n))$, we have that the j -th column of $\Phi\mathbf{P}$ has the form $(\Phi\mathbf{P})_j = [\phi(1)P_{j1} \phi(2)P_{j2} \dots \phi(n)P_{jn}]^T$, that is, $\sum_{i=1}^n (\Phi\mathbf{P})_j = \phi(i)$. Since $(\Phi\mathbf{P})^T = \mathbf{P}^T\Phi$ the i -th row of $\mathbf{P}^T\Phi$ is identical to the j -th column of $\Phi\mathbf{P}$ and thus $\sum_{j=1}^n (\mathbf{P}^T\Phi)^i = \phi(i)$. Consequently, the matrix $\Phi\mathbf{P} + \mathbf{P}^T\Phi$ is also symmetric and the sum of the elements in the columns in i -th row (or the sum of the elements in the rows in the same column) is like

$$\sum_{j=1}^n (\Phi\mathbf{P} + \mathbf{P}^T\Phi)^i = \sum_{j=1}^n (\Phi\mathbf{P} + \mathbf{P}^T\Phi)_i = \sum_{i,i \rightarrow j} \phi(i)P_{ij} + \underbrace{\sum_{j,j \rightarrow i} \phi(j)P_{ji}}_{\phi(i)}, \quad (1)$$

i.e. the sum of both incoming and outgoing probabilities. Since ϕ corresponds to the stationary distribution we have that $\sum_{i,i \rightarrow j} \phi(i)P_{ij} = \phi(i)$ for $(\Phi\mathbf{P})^T = \mathbf{P}^T\Phi = \phi^T$. Consequently, $\sum_{j=1}^n (\Phi\mathbf{P} + \mathbf{P}^T\Phi)^i = 2\phi(i) \forall i$. This leads to the definition of the following matrices:

$$\mathbf{L} = \Phi - \frac{\Phi\mathbf{P} + \mathbf{P}^T\Phi}{2} \quad \text{and} \quad \mathcal{L} = I - \frac{\Phi^{1/2}\mathbf{P}\Phi^{-1/2} + \Phi^{-1/2}\mathbf{P}^T\Phi^{1/2}}{2}, \quad (2)$$

where $\Phi = \text{diag}(\phi(1) \dots \phi(n))$, \mathbf{L} is the *combinatorial directed Laplacian* and \mathcal{L} is the *normalized directed Laplacian* [14]. Focusing on \mathbf{L} we have

$$L_{ij} = \begin{cases} \phi(i) & \text{if } i = j \\ -\left(\frac{\phi(i)P_{ij} + \phi(j)P_{ji}}{2}\right) & \text{otherwise.} \end{cases}, \quad (3)$$

where it is assumed that $P_{ii} = 0 \forall i$. Otherwise $L_{ii} = \phi(i)(1 - P_{ii})$.

Symmetrizing \mathbf{P} leads to real valued eigenvalues and eigenvectors. In addition Φ plays the role of a degree matrix and off-diagonal entries are designed so that the all-ones vector $\mathbf{1}$ is the eigenvector f_1 of the combinatorial Laplacian (the vector with eigenvalue $\lambda_1 = 0$). This is due to the fact that the sum of the i -th row of \mathbf{L} is $\sum_{j=1}^n (\mathbf{L})^i = \phi(i) - 2\phi(i)/2 = 0$. In any case, satisfying irreducibility is difficult in practice since sink vertices may arise frequently. For instance, a circular graph C_n given by $1 \rightarrow 2 \rightarrow 3 \rightarrow \dots \rightarrow n \rightarrow 1$ is clearly irreducible. However, the linear graph L_n obtained by removing $n \rightarrow 1$ from the cycle is reducible since we have a sink at n and the graph is no longer strongly connected. Sink vertices introduce rows of zeros in \mathbf{A} and consequently in \mathbf{P} . The consequence is the non-existence of a left Perron eigenvector and this renders computing the Laplacians is impossible. A formal trick consists of replacing \mathbf{P} by \mathbf{P}' so that $P'_{ij} = \frac{1}{n}$ if $A_{ij} = 0$ and $D_{ii} = 0$. This strategy is adopted in Pagerank [13] and allows for *teleporting* acting on the random walk to any other node in the graph. Teleporting is modeled by redefining \mathbf{P} in the following way: $\mathbf{P} = \eta\mathbf{P}' + (1 - \eta)\frac{\mathbf{1}\mathbf{1}^T}{n}$ with $0 < \eta < 1$. The new \mathbf{P} ensures both irreducibility and aperiodicity and this allows us to both apply \mathbf{P}' with probability η and to teleport from any node with $A_{ij} = 0$ with probability $1 - \eta$. In [15] a trade-off between large values η (preserving more the structure of \mathbf{P}') and small ones (potentially increasing the spectral gap) is recommended. For instance, in [16],

where the task is to learn classifiers on directed graphs, the setting is $\eta = 0.99$. When using the new \mathbf{P} we always have that $P_{ii} \neq 0$ due to the Pagerank masking. Such masking may introduce significant interferences in heat diffusion when the Laplacian is used to derive the heat kernel.

3 Directed Heat Kernels and Heat Flow

The definition of \mathbf{P} is critical for finding both the directed combinatorial Laplacian \mathbf{L} and the directed normalized Laplacian \mathcal{L} . Consequently it is also critical in determining the behavior of the heat kernel derived from the latter matrices. If the graph is strongly connected and aperiodic the original \mathbf{P} has a unique equilibrium distribution and the components of the combinatorial Laplacian are given by Eq. 3. Otherwise the above conditions are enforced by exploiting the Pagerank transformation. In any case, the $n \times n$ *heat/diffusion kernel* $\mathbf{K}_\beta(G)$ of the graph is the solution to the heat/diffusion equation: $\frac{\partial \mathbf{K}_\beta}{\partial \beta} = -\mathbf{L}\mathbf{K}_\beta$, and is given by the matrix exponentiation $\mathbf{K}_\beta(G) = \exp(-\beta\mathbf{L})$, for $\beta \geq 0$. Using the Taylor series (which may be useful for large n) we have: $\mathbf{K}_\beta(G) = e^{-\beta\mathbf{L}} = \mathbf{I}_n - \beta\mathbf{L} + \frac{\beta^2}{2!}\mathbf{L}^2 - \frac{\beta^3}{3!}\mathbf{L}^3 + \dots$, where \mathbf{I}_n is the $n \times n$ identity matrix. In this regard, the matrix $\mathbf{W} = \frac{\Phi\mathbf{P} + \mathbf{P}^T\Phi}{2}$ can be seen as the *weight matrix* of the undirected graph G_u associated with G (which may be also weighted) through \mathbf{P} and Φ . Therefore, the analysis of how the heat flows through G is equivalent to the analysis of how it flows through G_u .

We commence by reviewing the concept of *heat flow* [11]. Firstly, the spectral decomposition of the diffusion kernel is $\mathbf{K}_\beta(G) = \exp(-\beta\mathcal{L}) \equiv \Psi\Lambda\Psi^T$, where $\Lambda = \text{diag}(e^{-\beta\lambda_1}, e^{-\beta\lambda_2}, \dots, e^{-\beta\lambda_n})$, $\Psi = [\psi_1, \psi_2, \dots, \psi_n]$, and $\{(\lambda_i, \psi_i)\}_{i=1}^n$ are the eigenvalue-eigenvector pairs of $\Phi - \mathbf{W}$. Hence $K_{\beta_{ij}} = \sum_{k=1}^n \psi_k(i)\psi_k(j)e^{-\lambda_k\beta}$, and $K_{\beta_{ij}} \in [0, 1]$ is the (i, j) entry of a doubly stochastic matrix. Doubly stochasticity for all β implies *heat conservation* in the system as a whole. That is, not only in the nodes and edges of the graph but also in the *transitivity links* eventually established between non-adjacent nodes (if i is not adjacent to j , eventually will appear an entry $K_{\beta_{ij}} > 0$ for β large enough). The total *directed heat* flowing through the graph at a given β (*instantaneous directed flow*) is given by

$$F_\beta(G) = \sum_{i \rightarrow j} A_{ij} \left(\sum_{k=1}^n \psi_k(i)\psi_k(j)e^{-\lambda_k\beta} \right), \quad (4)$$

A more compact definition of the flow is $F_\beta(G) = \mathbf{A} : \mathbf{K}_\beta$, where $\mathbf{X} : \mathbf{Z} = \sum_{ij} X_{ij}Z_{ij} = \text{trace}(\mathbf{X}\mathbf{Z}^T)$ is the Frobenius inner product. While instantaneous flow for the heat flowing through the edges of the graph, it accounts neither for the heat remaining in the nodes nor for that in the transitivity links. The limiting cases are $F_0 = 0$ and $F_{\beta_{max}} = \frac{1}{n} \sum_{i \rightarrow j} A_{ij}$ which is reduced to $\frac{|E|}{n}$ if G is unattributed ($A_{ij} \in \{0, 1\} \forall i, j$). Defining F_β in terms of \mathbf{A} instead of \mathbf{W} , we retain the *directed* nature of the original graph G . The function derived from computing $F_\beta(G)$ from $\beta = 0$ to β_{max} is the so called *directed heat flow*

trace. These traces exhibit the following differences with respect to those of unattributed undirected graphs:

1. They satisfy the *phase transition principle* [11] (although the formal proof is out of the scope of this paper).
2. In general heat flow diffuses more slowly than in the undirected case and phase transition points (PTPs) appear later. This is due to the constraints imposed by \mathbf{A} .
3. PTPs may coincide with equilibrium points even when the directed graph is not the complete one. This happens in strongly connected graphs with many cycles (where connectivity constraints are relaxed) but the traces of single cycles do not have this behavior.
4. The sum of all walks connecting every pair of nodes is maximal (if we exclude the sum of all cycles for each node) for all components corresponding to non-zero values in \mathbf{A} . This is straightforward to prove by expressing the kernel in terms of sums of walks.
5. Graphs with at least one sink require the Pagerank mastering strategy which introduces noise in the diffusion process. This noise has no practical effect even for moderate/small values of η (e.g. $\eta = 0.15$).
6. The heat diffusion process does not only allow increasing heat values for setting transitivity links but it may also happen at directed edges. The main reason is that \mathbf{K}_β is expressed in terms of an undirected attributed graph given by \mathbf{W} even for non-attributed strongly connected and aperiodic digraphs.

4 Heat Flow - Thermodynamic Depth Complexity

The application of thermodynamic depth (TD) to characterize the complexity of directed graphs demands the formal specification of the micro-states whose history leads to the macro-state (of the network). Here we define such micro-states in terms of *expansion subgraphs*.

Let $G = (V, E)$ with $|V| = n$. Then the *directed history of a node* $i \in V$ is $h_i(G) = \{e(i), e^2(i), \dots, e^p(i)\}$ where: $e(i) \subseteq G$ is the *first-order expansion subgraph* given by i and all $j : i \rightarrow j$. If there are nodes j also satisfying $j \rightarrow i$ then these edges are included. If node i is a sink then $e(i) = i$. Similarly $e^2(i) = e(e(i)) \subseteq G$ is the *second-order expansion* consisting on $j \rightarrow z : j \in V_{e(i)}, z \notin V_{e(i)}$, including also $z \rightarrow j$ if these edges exists and $j \rightarrow z$. This process continues until p cannot be increased. If G is strongly connected $e^p(i) = G$, otherwise $e^p(i)$ is the strongly connected component to which i belongs.

Every $h_i(G)$ defines a different causal trajectory which may lead to G itself, if it is strongly connected, or to one of its strongly connected components otherwise. Thus, in terms of TD the full graph G or the union of its strongly connected components is the macro-state (macroscopic state). The *depth* of such macro-states relies on the variability of the causal trajectories leading to them. The higher the variability, the more complex it is to explain how the macro-state is reached and the deeper is this state. Therefore, in order to characterize

each trajectory we combine the heat flow complexities of its expansion subgraphs by means of defining *minimal enclosing Bregman balls* (MEBB) [18]. Here we use the I-Kullback-Leibler (I-KL) Bregman divergence between traces f and g : $D_F(\mathbf{f}||\mathbf{g}) = \sum_{i=1}^d f_i \log \frac{f_i}{g_i} - \sum_{i=1}^d f_i + \sum_{i=1}^d g_i$ with convex generator $F(\mathbf{f}) = \sum_{i=1}^d (f_i \log f_i - f_i)$.

Given $h_i(G)$, the heat flow complexity $\mathbf{f}_t = f(e^t(i))$ for the t -th expansion of i , a generator F and a Bregman divergence D_F , the *causal trajectory* leading to G (or one of its strongly connected components) from i is characterized by the center $\mathbf{c}_i \in R^d$ and radius $r_i \in R$ of the MEBB $\mathcal{B}^{\mathbf{c}_i, r_i} = \{\mathbf{f}_t \in \mathcal{X} : D_F(\mathbf{c}_i||\mathbf{f}_t) \leq r_i\}$. Solving for the center and radius implies finding \mathbf{c}^* and r^* minimizing r subject to $D_F(\mathbf{c}_i||\mathbf{f}_t) \leq r \forall t \in \mathcal{X}$ with $|\mathcal{X}| = T$. Considering the Lagrange multipliers α_t we have that $\mathbf{c}^* = \nabla^{-1}F(\sum_{t=1}^T \alpha_t \mathbf{f}_t \nabla F(\mathbf{f}_t))$. The efficient algorithm in [18] estimates both the center and multipliers. This idea is closely related to Core Vector Machines [19], and it is interesting to focus on the non-zero multipliers (and their support vectors) used to compute the optimal radius. More precisely, the multipliers define a convex combination and we have $\alpha_t \propto D_F(\mathbf{c}^*||\mathbf{f}_t)$, and the radius is simply chosen as: $r^* = \max_{\alpha_t > 0} D_F(\mathbf{c}^*||\mathbf{f}_t)$.

Given the directed graph $G = (V, E)$, with $|V| = n$ and all the n pairs (\mathbf{c}_i, r_i) , the *heat flow-thermodynamic depth complexity* of G is characterized by the MEBB $\mathcal{B}^{\mathbf{c}, r} = \{\mathbf{c}_i \in \mathcal{X}_i : D_F(\mathbf{c}||\mathbf{c}_i) \leq r\}$. As a result, the *TD depth of the directed graph* is given by $\mathcal{D}(G) = r$. This definition of depth is highly consistent with summarizing node histories with second-order variability operators to find a global causal trajectory which is as tightly bounded as possible.

5 Experiments: Analysis of Language Complexity

We analyse networks extracted from the adjacency of words for different languages. We used a subset of the parallel corpora published in the Official Journal of the European Union. We used 100,000 lines of text from each language, all of them corresponding to the same text (human translation). The languages included in this study are: Bulgarian (BG), Czech (CS), Danish (DA), German (DE), Greek (EL), English (EN), Spanish (ES), Estonian (ET), Finnish (FI), French (FR), Hungarian (HU), Italian (IT), Lithuanian (LT), Latvian (LV), Maltese (MT), Dutch (NL), Polish (PL), Portuguese (PT), Slovak (SK), Slovene (SL) and Swedish (SV).

The directed adjacency graph represent words which appear consecutively in a text. We take the words as they appear in the text (surface form) and not only their lemmas. In this way we retain morphology, which imposes different restrictions in each language. In the graphs we construct the edges commencing from each node (word) V_i connect to the words which follow V_i . Thus a language with no restrictions is represented by a fully connected graph. We also take into account the frequency of each connection occurring in the corpus, and we store this information as attributes for the directed edges. This means that we give a greater importance to those adjacencies between words which are used more frequently (in the corpus). Although we do not store the frequency of each word

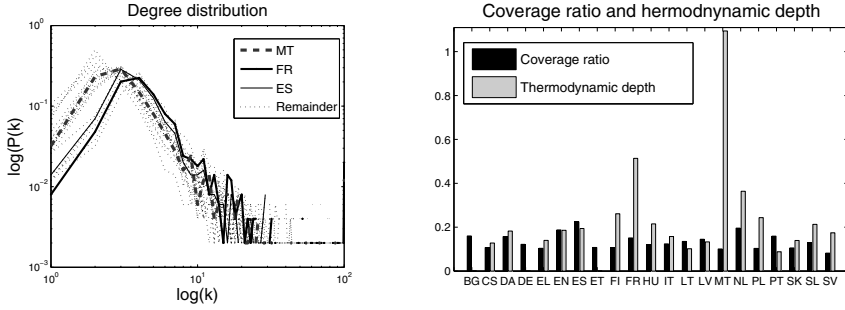


Fig. 1. Left: Log-log plot of the degree distribution of three languages (MT, FR, ES). The remaining languages are represented as well. All distributions behave in a similar way. We found no significant difference between the degree distributions of different languages. Right: The coverage measures the amount of text covered by the n most frequent surface forms.

in the graph representation we do use it for selecting the n most frequent words which constitute the nodes V_i .

In Fig. 1-left we show that the classical analysis based on the degree distribution is blind to differences of complexities between languages (all languages follow a similar degree distribution). In Fig. 1-right we compare the thermodynamic depth for different languages and show the amount of text that was covered by the graph of each language. Each of the graphs have $n = 500$ nodes which correspond to the n most frequent surface forms. These n surfaces cover part of the corpus of the language and the remainder of the surfaces in the corpus are not represented by the graph because of their lower frequency. We may take as a baseline for the complexity of a language the coverage ratio of n surface forms. An intuitive explanation is that if all the languages had a similar number of different lemmas in the parallel corpora, then the number of different surface forms would depend on the morphology of each language. A simpler morphology would enable the n surfaces to cover a larger amount of text than that covered by a rich morphology. This baseline does not capture all the subtle complexities of the network formed by the adjacency relation between words. The bar plot shows that there are some languages which do have the same tendencies both for thermodynamic depth and coverage. This is not the case of FR and MT.

6 Conclusions and Future Work

In this work we extend the heat-diffusion TD initially designed for unattributed undirected graphs to digraphs. We analyze the Laplacian operator used to that end and the consequences of using it to compute the heat flow. We enunciate several properties of heat diffusion traces in digraphs. In our experiments we compute the complexity of several languages and find differences that are blind to degree distribution analysis. Our future work includes the formal proof of the properties and the exploration of other graph-based representations of languages.

Acknowledgements. Francisco Escolano was funded by project TIN2011-27043 of the Spanish Government. Edwin Hancock was supported by a Royal Society Wolfson Research Merit Award.

References

1. Escolano, F., Suau, P., Bonev, B.: Information Theory in Computer Vision and Pattern Recognition. Springer, London (2009)
2. Torsello, A., Hancock, E.R.: Learning Shape-Classes Using a Mixture of Tree-Unions. *IEEE Tran. on Pattern Analysis and Mach. Intelligence* 28(6), 954–967 (2006)
3. Torsello, A., Lowe, D.L.: Supervised Learning of a Generative Model for Edge-Weighted Graphs. In: *Proc. of ICPR* (2008)
4. Körner, J.: Coding of an Information Source Having Ambiguous Alphabet and the Entropy of Graphs. In: *Transactions of the 6th Prague Conference on Information Theory*, pp. 411–425 (1973)
5. Passerini, F., Severini, S.: The von Neumann Entropy of Networks. [arXiv:0812.2597v1](https://arxiv.org/abs/0812.2597v1) (December 2008)
6. Han, L., Escolano, F., Hancock, E.R., Wilson, R.: Graph Characterizations From Von Neumann Entropy. *Pattern Recognition Letters* (in press, 2012)
7. Feldman, D.P., Crutchfield, J.P.: Measures of Statistical Complexity: Why? *Physics Letters A* 238(4-5), 244–252 (1998)
8. Estrada, E.: Quantifying Network Heterogeneity. *Phys. Rev. E* 82, 066102 (2010)
9. Randić, M.: Characterization of Molecular Branching. *Journal of the American Chemical Society* 97(23), 6609–6615 (1975)
10. Lloyd, S., Pagels, H.: Complexity as Thermodynamic Depth *Ann. Phys.* 188, 186 (1988)
11. Escolano, F., Hancock, E.R., Lozano, M.A.: Heat Diffusion: Thermodynamic Depth Complexity of Networks. *Phys. Rev. E* 85, 036206 (2012)
12. Escolano, F., Lozano, M.A., Hancock, E.R., Giorgi, D.: What Is the Complexity of a Network? The Heat Flow-Thermodynamic Depth Approach. In: Hancock, E.R., Wilson, R.C., Windeatt, T., Ulusoy, I., Escolano, F. (eds.) *SSPR&SPR 2010*. LNCS, vol. 6218, pp. 286–295. Springer, Heidelberg (2010)
13. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank Citation Ranking: Bring Order to the Web (Technical Report). Stanford University (1998)
14. Chung, F.: Laplacians and the Cheeger Inequality for Directed Graphs. *Annals of Combinatorics* 9, 1–19 (2005)
15. Johns, J., Mahadevan, S.: Constructing Basic Functions from Directed Graphs for Value Functions Approximation. In: *Proc. of ICML* (2007)
16. Zhou, D., Huang, J., Schölkopf, B.: Learning from Labeled and Unlabeled Data on a Directed Graph. In: *Proc. of ICML* (2005)
17. Brasseur, C.E., Grady, R.E., Prassidis, S.: Coverings, Laplacians and Heat Kernels of Directed Graphs. *Electr. J. Comb* 01/2009 16(1) (2009)
18. Nock, R., Nielsen, F.: Fitting Smallest Enclosing Bregman Ball. In: Gama, J., Camacho, R., Brazdil, P.B., Jorge, A.M., Torgo, L. (eds.) *ECML 2005*. LNCS (LNAI), vol. 3720, pp. 649–656. Springer, Heidelberg (2005)
19. Tsang, I.W., Kocsor, A., Kwok, J.T.: Simple Core Vector Machines with Enclosing Balls. In: *Proc. of ICLM* (2007)