Complex networks classification using the deterministic tourist walk

Oliveira, Maria Vithória - (NUSP: 15420024)[a], Travieso, G.[a], Peron, T.[b]

[a]IFSC, USP, São Carlos, SP, Brazil
[b]ICMC, USP, São Carlos, SP, Brazil

Abstract

The classification of complex networks is vital for understanding their structural and functional diversity, enabling the application of specific analytical methods and the development of theoretical models. In addition, classifying networks is essential to predict dynamic behavior and implement control strategies, such as containing epidemics or restoring neural functions. Developing effective classification methods in complex networks is challenging. Traditional methods often fail to define relationships and patterns, so researchers use machine learning and statistical methods to categorize networks by their structures. This paper classifies synthetic networks using the Deterministic Tourist Walk (DTW) algorithm and Linear Discriminant Analysis (LDA) on trajectory histograms. Synthetic networks were generated by four models: Barabasi-Albert, Watts-Strogatz, Erds-Rényi, and Geographical. The study compared two walking rules in DTW to assess the performance of network classification. Findings showed that using two types of network exploration rules yielded a classification accuracy of 98.33%, which was more efficient than only comparing multiple memory sizes.

Keywords: Deterministic Tourist Walk, Complex Networks Classification, Walking Rules.

1. Introduction

A complex network is a system composed of a large number of interconnected elements, where the interactions and structural patterns within the network give rise to behaviors that cannot be understood by analyzing individual components independently. These behaviors emerge only when the system is considered as a whole [1, 2, 3, 4]. The application of networks and graph theory to studying real-world issues began in 1736, with Leonhard Euler employing graph theory to address the seven bridges of the Königsberg problem [5]. Complex networks are now used in various scientific fields, such as the physical, biological, and social sciences, to model and analyze systems with relationships and dependencies [1, 3, 4].

One of the challenges in the complex networks field is the development of effective classification methods [6]. Traditional methods often fail to capture the definition of relationships and patterns. Consequently, researchers employ various approaches, including machine learning and statistical methods, to categorize networks based on their structural properties. Nevertheless, classification remains difficult due to the diverse topologies, noise, and dynamic nature of real-world systems [2].

Walk-based methods, such as the Deterministic Tourist Walk (DTW), offer an alternative approach to network classification: Through a systematic examination of the network and the extraction of information from the paths taken by agents, it becomes possible to discover structural features that can be used for classification [6].

Initially proposed for texture analysis in images, the DTW method achieved notable success by capturing subtle patterns that are often missed by traditional methods. Using a systematic exploration of the image structures, DTW effectively extracted texture signatures [7, 8], distinguishing between textures with high accuracy, showcasing its robust potential for analyzing structured data [9].

This paper aims to classify synthetic complex networks using the Deterministic Tourist Walk, employing the Linear Discriminant Analysis (LDA) application of trajectory histograms [10, 11, 12]. Two walking rules will be used in the DTW algorithm to allow for a comparative evaluation of the results, the selection of the next nodes with:

- the minimal degree difference;
- the maximum degree difference between nodes.

The paper is organized as follows: Section 2 provides an overview of the Deterministic Tourist Walk method. Section 3 presents the details of the work methodology. Section 4 describes the experimental results and discussion. Finally, Section 5 concludes the paper.

2. Overview

Deterministic Tourist Walk (DTW), developed as an agent-based approach in the early twentieth century, originally aimed at extracting texture signatures from images [13, 6]. Despite its initial purpose in texture analysis, DTW is a versatile method, capable of being applied to networks as well. The DTW method is a deterministic automaton moving through a system, similar to a random walk but governed by specific walking rules. The term Deterministic Tourist Walk originates from its analogy to a tourist systematically exploring a city map to optimize a path based on measures such as minimal distance or cost [13, 9].

In the DTW approach, each agent starts at a node within a graph and moves to the subsequent node, one at a time. The memory size (μ) parameter is configured for the tourist, thereby prohibiting the return to cities (nodes) already stored in their memory. With a memory size of 0, the tourist can visit any city without restrictions. If $\mu = 1$ the tourist can remember only the last city visited [13, 6, 7].

The tourist trajectory is divided into two distinct phases: the transient phase, which has a length of t, and the attractor phase, attractive enough to trap the tourist, ending the trajectory, as shown in Figure 2. According to the constraints imposed by the memory, the attractor is defined to have a length of $a = \mu + 1$. Furthermore, the total trajectory is given by the sum of the transient and attractor segments: $l = t + a$.

Starting the tourist walk from each distinct city on the map produces n trajectories, each with distinct combinations of transient attractors, creating a histogram $h(\ell)$ of the collective distribution. For a trajectory with length ℓ, the histogram is given by

$$h(\ell) = \sum_{b=0}^{\ell-1} S(b, \ell - b) \tag{1}$$

with $S(t, p)$ being the counting function,

$$S(t, p) = \frac{1}{n} \sum_{i=1}^{n} \begin{cases} 1, & \text{if } t = \tau_i \text{ and } p = P_i \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where τ_i is the *ith* transient and P_i is the *ith* attractor and n is the number of trajectories [6].

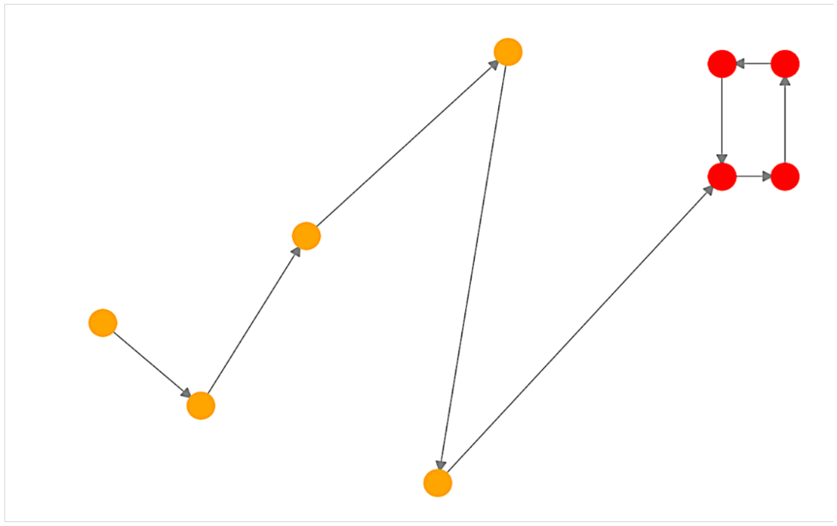

Figure 1: The orange nodes represent the transient trajectory and the red nodes depict the attractor trajectory, where the tourist is trapped.

The signature vector Ψ for a given rule r and for a memory size μ can be derived from the transient versus attractor histogram:

$$\Psi_{\mu}^{r} = [h(\ell = \mu + 1), h(\ell = \mu + 2), \dots, h(\ell = \mu + z)] \tag{3}$$

where z is arbitrarily defined.

Backes et al. [7] demonstrated that the performance of the DTW method can be improved by concatenating the vector from Equation 3 into different memory values, thus creating a new vector,

$$\Psi_{\vec{\mu}}^{r} = \psi_{\vec{\mu}_1}^{r}, \psi_{\vec{\mu}_2}^{r}, \dots, \psi_{\vec{\mu}_k}^{r}. \tag{4}$$

in which $\vec{\mu} = [\mu_1, \mu_2, \dots, \mu_k]$ is a vector of memory sizes.

Another way to enhance the performance of DTW is to concatenate the vector from Equation 5 for several walking rules [8, 7],

$$\Psi_{\vec{\mu}}^{\vec{r}} = \psi_{\vec{\mu}}^{r_1}, \psi_{\vec{\mu}}^{r_2}, \dots, \psi_{\vec{\mu}}^{r_j} \tag{5}$$

wherein $\vec{r} = [r_1, r_2, \dots, r_j]$ is the vector of walking rules.

With walking rules, the tourist chooses the next node to visit. These rules can be based on network properties; however, the most commonly used rules include selecting the node with the highest and lowest degree, or the minimal and maximum degree difference between nodes.

The DTW has three stop conditions:

- if the tourist meets an attractor, the tourist stops;
- if the tourist finds a city at the extreme end of the map, i.e. the only neighbor is the city from the previous step and, due to memory constraints, the tourist cannot return to it;
- if the tourist exceeds a predefined maximum number of steps (typically $\mu + z$), the tourist stops.

3. Methodology

This paper aims to classify synthetic complex networks using the Deterministic Tourist Walk Algorithm. The dataset consists of synthetic networks generated by four models: Barábasi-Albert, Watts-Strogatz, Erdős–Rényi and Geographical network. All experiments were conducted using 50 networks with 500 nodes and an average degree $k = 4$.

The study of DTW method applied two different walking rules to compare their performance in network classification. The rules considered are:

- The tourist chooses the next node with the minimal degree difference from the current node;
- The tourist chooses the next node with the maximum degree difference from the current node.

Histograms of transients *versus* attractors were generated using walking rules. Thus, to classify synthetic complex networks, Linear Discriminant Analysis (LDA) was applied.

3.1. Classification Method

The Linear Discriminant Analysis (LDA) is a statistical technique widely used for dimensionality reduction and supervised classification. LDA projects data onto a new space with the aim of maximizing the separation between classes while minimizing variability within the same class. This approach is particularly useful in scenarios where classes are linearly or nearly linearly separable. The method assumes that the data for each class follow a Gaussian distribution with the same covariance matrix, which simplifies the model and ensures computational efficiency [10, 11, 12].

In this study, LDA is used to classify the synthetic networks generated by the Barabási-Albert, Watts-Strogatz, Erdős-Rényi, and Geographical network models. By applying LDA to the signature vectors obtained through the Deterministic Tourist Walk (DTW) algorithm, the data is projected into a space that maximizes the separation between different network classes.

4. Results and Discussion

Numerical simulations were performed in order to classify the synthetic complex networks into Watts-Strogatz, Erdős-Rényi, Barabási-Albert and Geographical networks using the Deterministic Tourist Walk method. In addition, histograms were constructed to compare the transient and attractor walk distributions.

4.1. The transient vs. attractor histogram

The two-dimensional histogram of transient vs. attractor shows the influence of the μ parameter on tourist dynamics, as depicted in Figure 2. For $\mu = 1$, the distribution of the attractors in both the minimal and maximum degree difference rules was highly concentrated, with the attractors mostly fixed at short lengths, mainly around 2. In contrast, the transients exhibited lower concentrations with minimal dispersion, indicating that the system quickly stabilized into simple recurring cycles.

Upon increasing the parameter $\mu = 4$, the variety of attractor lengths increased significantly. Within the framework of the minimal degree difference rule, the attractors exhibited a range of 2 to 10, indicating that the increased parameter μ enhanced the exploration of longer and more complex cycles. Additionally, the maximum degree difference rule demonstrated even greater variability, with attractor lengths reaching up to 12 and displaying a higher density in extended cycles. This suggests that the increase in the parameter μ promotes the investigation of prolonged recurrent states, particularly under the maximum degree difference rule, which tends to traverse broader trajectories.

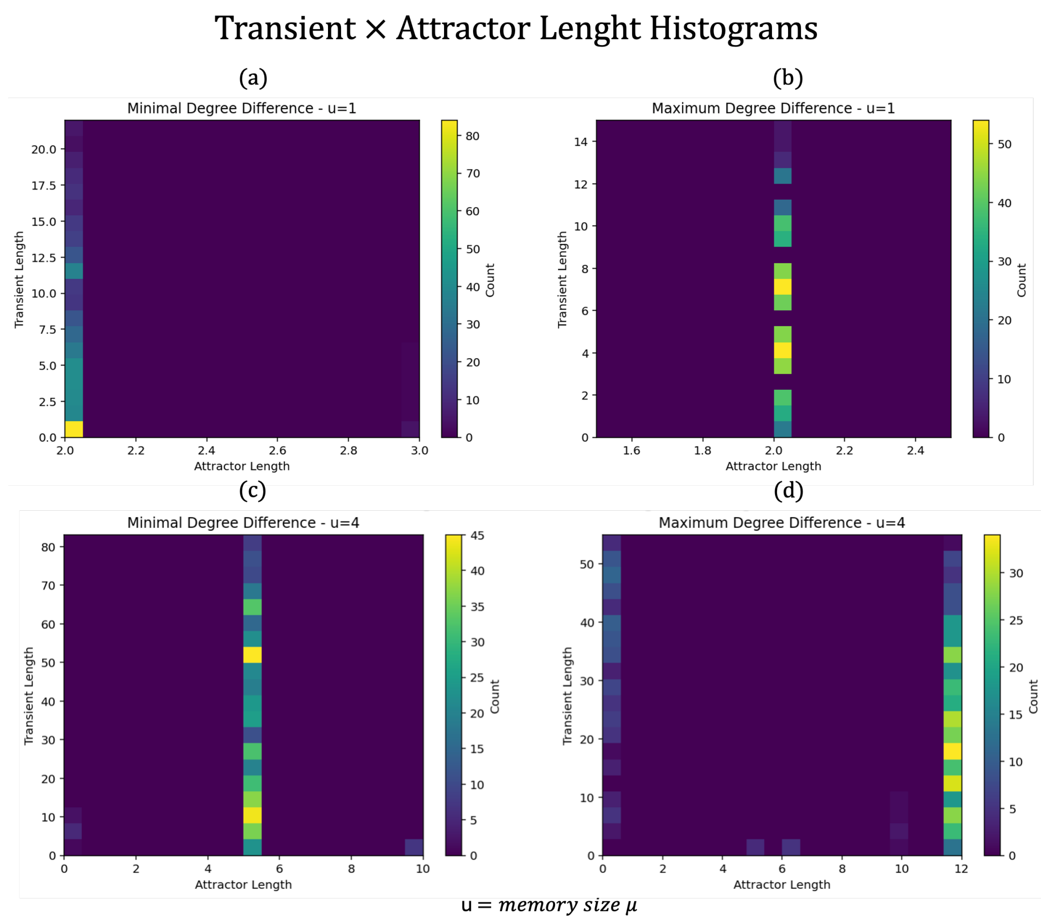

Figure 2: Comparison of transient lengths and attractor lengths for both rules is depicted in the histograms. Figures (a) and (c) employ the minimal degree difference rule, whereas figures (b) and (d) apply the maximum degree difference rule. Figures (a) and (b) utilize a memory size $\mu = 1$, while figures (c) and (d) incorporate a memory size $\mu = 4$.

4.2. Classification results

The signature vector was concatenated with the memory sizes: (i) $\mu = [1, 2, 3]$, (ii) $\mu = [1, 2, 3, 4]$ and (iii) $\mu = [1, 2, 3, 4, 5]$. In an effort to enhance the performance of the complex network classifier, the walking rules were also incorporated through concatenation, thereby integrating both the minimum and maximum degree difference walking rules relating to the network's nodes.

Table 1 shows a comparison of the accuracy of the two rules applied. (a), (b) and (c) correspond to the minimal degree difference rule, while (d), (e), and (f) relate to the maximum degree difference rule. The analysis reveals that the application of the minimal degree difference rule yields greater accuracy in classifying synthetic networks compared to the maximum degree difference rule.

Minimal degree difference rule		**Maximum degree difference rule**	
Signature vector	**Accuracy (%)**	**Signature vector**	**Accuracy (%)**
(a) $\Psi_{\vec{\mu}}^{r}[1,2,3]$	71.67	(d) $\Psi_{\vec{\mu}}^{r}[1,2,3]$	41.67
(b) $\Psi_{\vec{\mu}}^{r}[1,2,3,4]$	76.67	(e) $\Psi_{\vec{\mu}}^{r}[1,2,3,4]$	66.67
(c) $\Psi_{\vec{\mu}}^{r}[1,2,3,4,5]$	81.67	(f) $\Psi_{\vec{\mu}}^{r}[1,2,3,4,5]$	53.33

Table 1: Accuracy of the DTW method using two walking rules applied to synthetic networks.

Backes et al. [8, 7] have demonstrated that the integration of two walking rules enhances accuracy, as shown in Table 2. Moreover, result (i) in Table 2 suggests that an increase in memory size in the signature vector improves the precision of the complex network classifier, achieving an accuracy of 98.33%.

Signature Vector	**Accuracy (%)**
(g) $\Psi_{\vec{\mu}}^{r}[1,2,3]$	95.00
(h) $\Psi_{\vec{\mu}}^{r}[1,2,3,4]$	95.00
(i) $\Psi_{\vec{\mu}}^{r}[1,2,3,4,5]$	98.33

Table 2: Accuracy of the DTW method using a combination of the two walking rules in the signature vector, applied to synthetic networks.

4.3. Discussion

The signature vector was concatenated on the basis of memory sizes and walking rules. The findings indicated that applying two distinct types of vector concatenation produced a classification accuracy of 98.33%, demonstrating greater efficiency than employing only the concatenation of the memory size vector. Moreover, the transient vs. attractor histograms demonstrate that the increase in the μ parameter improved the exploration of longer and more complex cycles.

5. Conclusions

This study classified synthetic complex networks using the Deterministic Tourist Walk (DTW) algorithm, employing the Linear Discriminant Analysis (LDA) application of trajectory histograms. For such aims, numerical simulations were conducted to classify the complex networks into Watts-Strogatz, Erdős-Rényi, Barabási-Albert and Geographical networks.

The study applied two different walking rules to compare their performance in network classification. The findings indicated that applying two distinct types of rules in DTW resulted in a classification accuracy of 98.33%, demonstrating greater efficiency than using only multiple memory sizes.

These results explain how the minimal and maximum degree difference rules affect the tourist's journey on the networks' classification. Future studies should explore more walking rules, their combinations, and test them on various distinct networks.

References

[1] S. H. Strogatz, Exploring complex networks, Nature 410 (2001) 268–276. URL: http://dx.doi.org/10.1038/35065725. doi:10.1038/35065725.

[2] B. Kantarci, V. Labatut, Classification of complex networks based on topological properties, in: 2013 International Conference on Cloud and Green Computing, IEEE, 2013, p. 297–304. URL: http://dx.doi.org/10.1109/CGC.2013.54. doi:10.1109/cgc.2013.54.

[3] T. HASEGAWA, An introduction to complex networks, Interdisciplinary Information Sciences 17 (2011) 175–195. URL: http://dx.doi.org/10.4036/iis.2011.175. doi:10.4036/iis.2011.175.

[4] E. Estrada, Introduction to complex networks: structure and dynamics, in: Evolutionary equations with applications in natural sciences, Springer, 2014, pp. 93–131.

[5] G. Caldarelli, Complex networks: Principles, methods and applications by vito latora, vincenzo nicosia and giovanni russo, Journal of Complex Networks 6 (2018) 830–830. URL: http://dx.doi.org/10.1093/comnet/cnx062. doi:10.1093/comnet/cnx062.

[6] W. N. Gonçalves, A. S. Martinez, O. M. Bruno, Complex network classification using partially self-avoiding deterministic walks, Chaos: An Interdisciplinary Journal of Nonlinear Science 22 (2012). URL: http://dx.doi.org/10.1063/1.4737515. doi:10.1063/1.4737515.

[7] A. R. Backes, W. N. Gonçalves, A. S. Martinez, O. M. Bruno, Texture analysis and classification using deterministic tourist walk, Pattern Recognition 43 (2010) 685–694. URL: http://dx.doi.org/10.1016/j.patcog.2009.07.017. doi:10.1016/j.patcog.2009.07.017.

[8] A. R. Backes, O. M. Bruno, M. G. Campiteli, A. S. Martinez, Deterministic tourist walks as an image analysis methodology based, Progress in Pattern Recognition Image Analysis and Applications (2006) 784–793.

[9] M. G. Campiteli, P. D. Batista, O. Kinouchi, A. S. Martinez, Deterministic walks as an algorithm of pattern recognition, Physical Review E 74 (2006). URL: http://dx.doi.org/10.1103/PhysRevE.74.026703. doi:10.1103/physreve.74.026703.

[10] Z.-P. Liu, Linear discriminant analysis, Encyclopedia of Systems Biology (2013) 1132–1133.

[11] A. Tharwat, T. Gaber, A. Ibrahim, A. E. Hassanien, Linear discriminant analysis: A detailed tutorial, AI Communications 30 (2017) 169–190. URL: http://dx.doi.org/10.3233/AIC-170729.

[12] S. Zhao, B. Zhang, J. Yang, J. Zhou, Y. Xu, Linear discriminant analysis, Nature Reviews Methods Primers 4 (2024). URL: http://dx.doi.org/10.1038/s43586-024-00346-y. doi:10.1038/s43586-024-00346-y.

[13] J. V. Merenda, O. M. Bruno, Using deterministic self-avoiding walks as a small-world metric on watts–strogatz networks, Physica A: Statistical Mechanics and its Applications 621 (2023) 128713. URL: http://dx.doi.org/10.1016/j.physa.2023.128713. doi:10.1016/j.physa.2023.128713.