

The background features a dark red color with a network of red dashed lines. At the top and bottom, there are large, white, stylized letters 'D' and 'S' respectively, which are partially cut off by the edges of the slide. The main title is centered in white text.

A non-linear approach for pattern recognition in networks

Marina Jeaneth Machicao Justo

Advisor: Prof. Dr. Odemir M. Bruno

18 de setembro de 2017

Large, white, stylized letters 'SIFSO' are positioned at the bottom of the slide, partially overlapping the red dashed lines and the dark red background.


Ph. D Project Profile

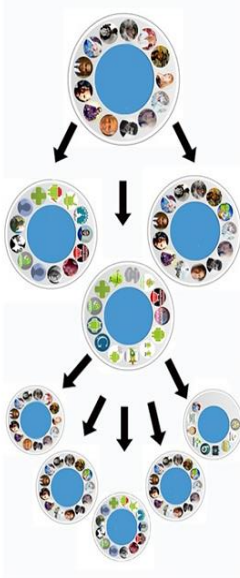
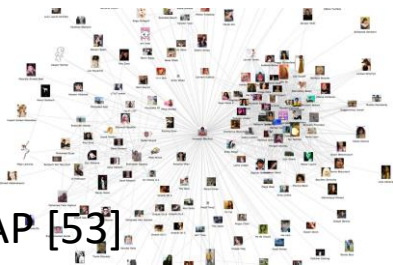
- Ph. D elapsed time: 4 years
- Expected Ph. D defense: December 2017
- Thesis deposit deadline: March 2018

Networks are everywhere!

Social networks

- Twitter 

- Google + 



Dataset SNAP [53]

Stomata distribution networks

Tradescantia zebrina

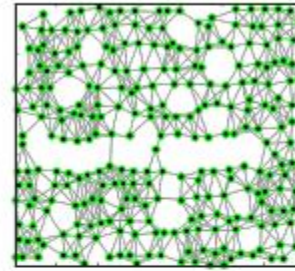
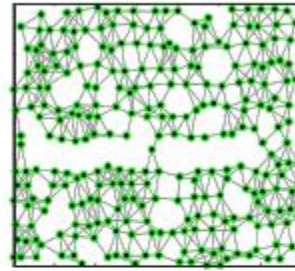
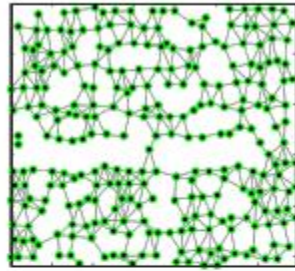
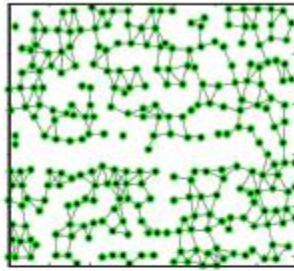
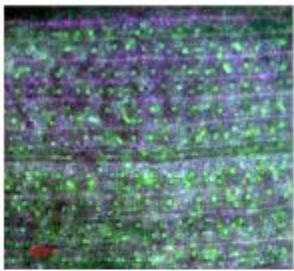
400 μm

460 μm

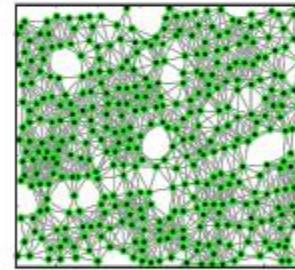
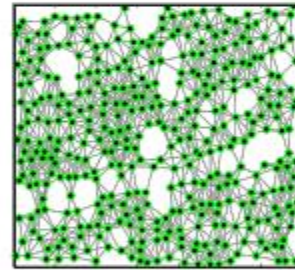
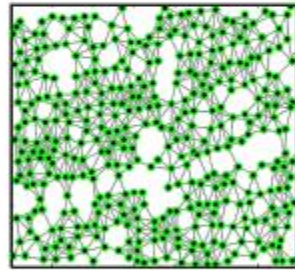
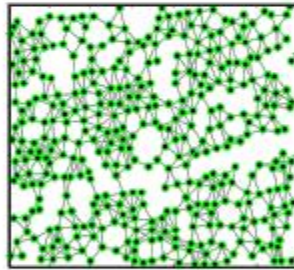
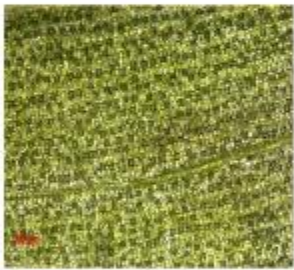
520 μm

580 μm

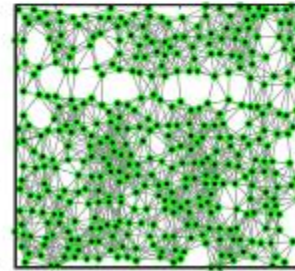
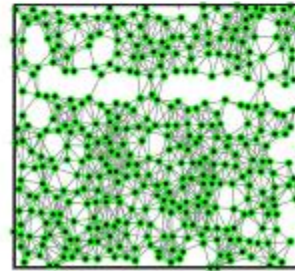
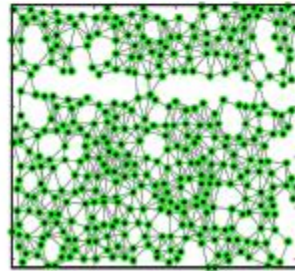
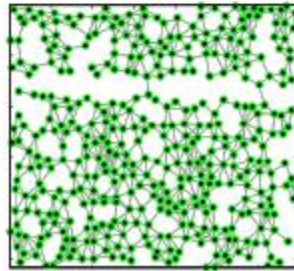
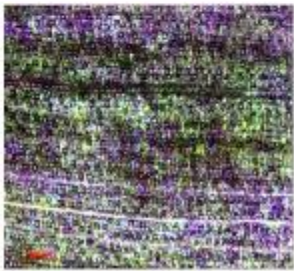
Exposição de Luz



natural



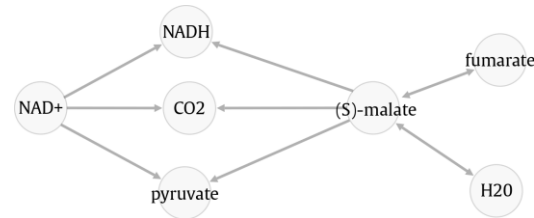
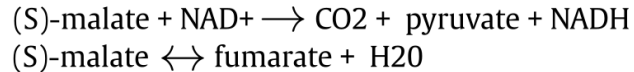
4h



24h

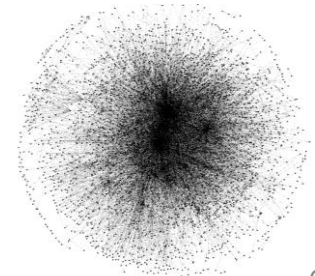
Metabolic networks

- Metabolites (nodes) relationship are connected according to the reaction directions.



- Organisms [18]:

- Archaea
- Bacterium
- Eukaryote



Textual networks

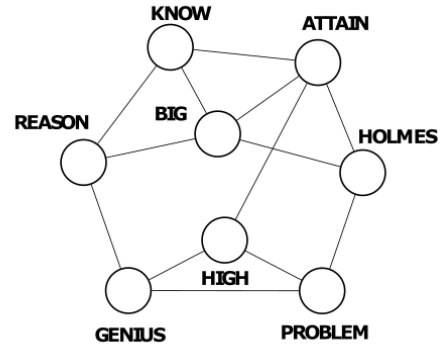
1) Pre-processing

1.1) Lemmatization

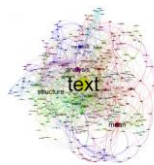
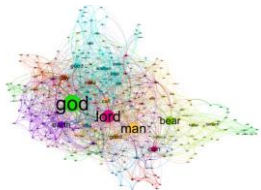
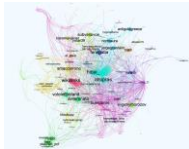
Twice already in his career had Holmes helped him to attain success, his own sole reward being the intellectual joy of the problem. For this reason the affection and respect of the Scotchman for his amateur colleague were profound, and he showed them by the frankness with which he consulted Holmes in every difficulty. Mediocrity knows nothing higher than itself, but talent instantly recognizes genius, and MacDonald had talent enough for his profession to enable him to perceive that there was no humiliation in seeking the assistance of one who already stood alone in Europe, both in his gifts and in his experience. Holmes was not prone to friendship, but he was tolerant of the big Scotchman, and smiled at the sight of him.

profession	nothing	intellectual
big	enable	talent
instantly	every	prone
know	seek	show
colleague	sight	smile
mediocrity	already	attain
sole	stand	holmes
assistance	reason	scotchman
friendship	joy	high
macdonald	frankness	respect
alone	twice	difficulty
problem	reward	perceive
enough	gift	europe
experience	career	profound
success	affection	one
consult	recognize	tolerant
help	genius	

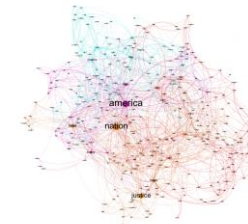
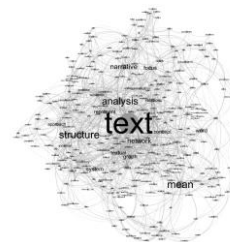
2) Network construction



Poe



Darwin



Pattern recognition in networks

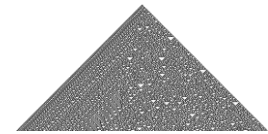
- Emerges due to high demand on big data scenario.
- Aims to characterize networks
 - Extracts information from the correlation between vertices and their relationship to the network.
- Literature:
 - Structural measurements [7],
 - Some attempts based on non-linear methods, such as random walks [10].

Cellular automata (artificial life)

- Discrete dynamical systems in time-space [3].
- Well-known CAs
 - ECA (Wolfram)
 - Life-Life (Conway)
- Represented by



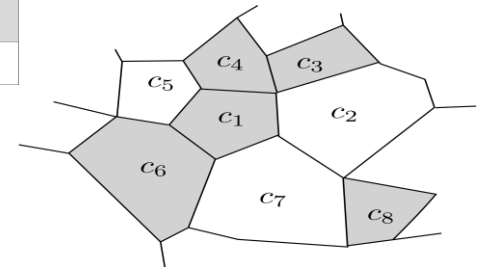
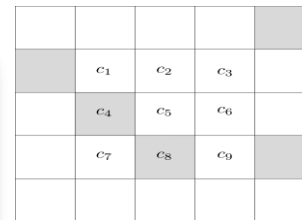
Specie *conus textile*



Rule 30 ECA

AC representada por $\langle \mathcal{T}, S, s, \mathcal{N}, \phi \rangle$

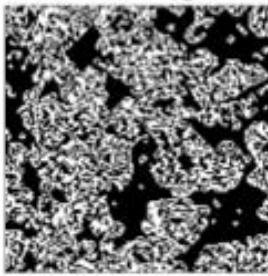
- \mathcal{T} : tesselação
- S : Estados, $S = \{s_0, s_1, \dots, s_{k-1}\}$
- s : função de estado, $s(c_i, t)$
- \mathcal{N} : função vizinhança
- ϕ : regra de transição



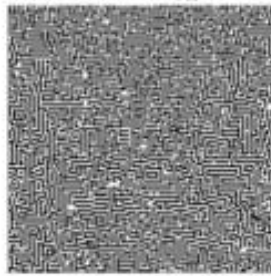
Life-like CA

- Nomenclature: Bx/Sy (2^{18} rules)

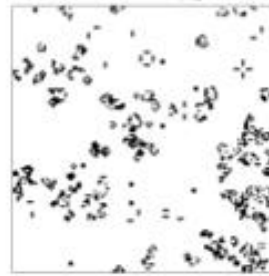
Coagulation
B378/S235678



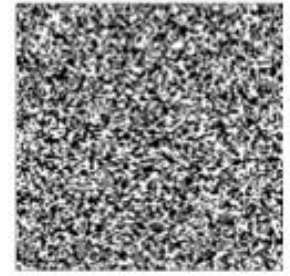
Mazetric
B3/S1234



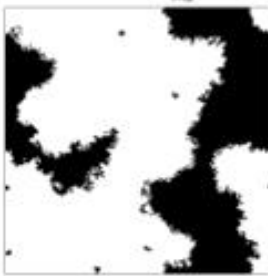
PseudoLife
B357/S238



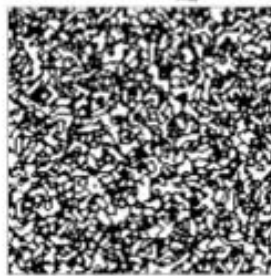
B1357/S2468



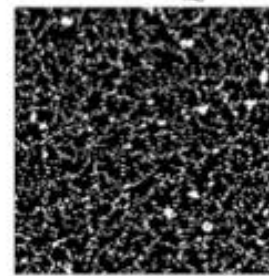
Day&Night
B3678/S34678



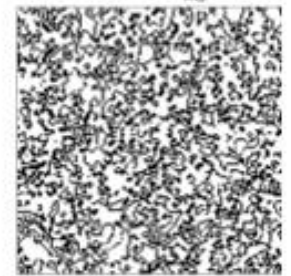
WalledCities
B45678/S2345



Coral
B3/S45678



Serviettes
B234/S



Cellular automata in networks

Cellular automata

- Entities: Cells
- Tessellation:
 - Regular/irregular
- Local neighborhood
- Rule
 - Based on #cells alive

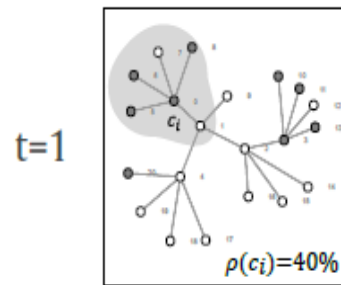
Network automata

- Nodes
- Tessellation:
 - Network
- Non-spatial neighborhood
 - not necessarily
- Rule:
 - Based on neighborhood density

$$s(c_i, t + 1) = \begin{cases} 1, & \text{if } s(c_i, t) = 0 \text{ and } x/r \leq \rho_i < (x + 1)/r \Rightarrow \text{born (B) rule} \\ 1, & \text{if } s(c_i, t) = 1 \text{ and } y/r \leq \rho_i < (y + 1)/r \Rightarrow \text{survive (S) rule} \\ 0, & \text{otherwise,} \end{cases}$$

Life-like network automata (LLNA)

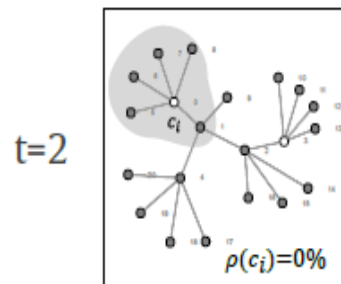
- Life-like family has 2^{18} rules (possible solutions)



Rule: B3/S23

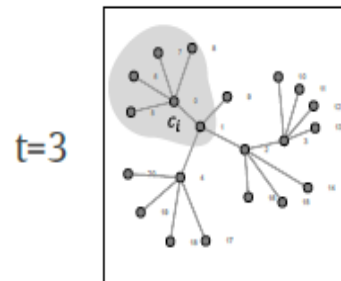
$s(c_i, 1) = 0$ and $33\% \leq \rho(c_i) < 44\%$

⇓ born



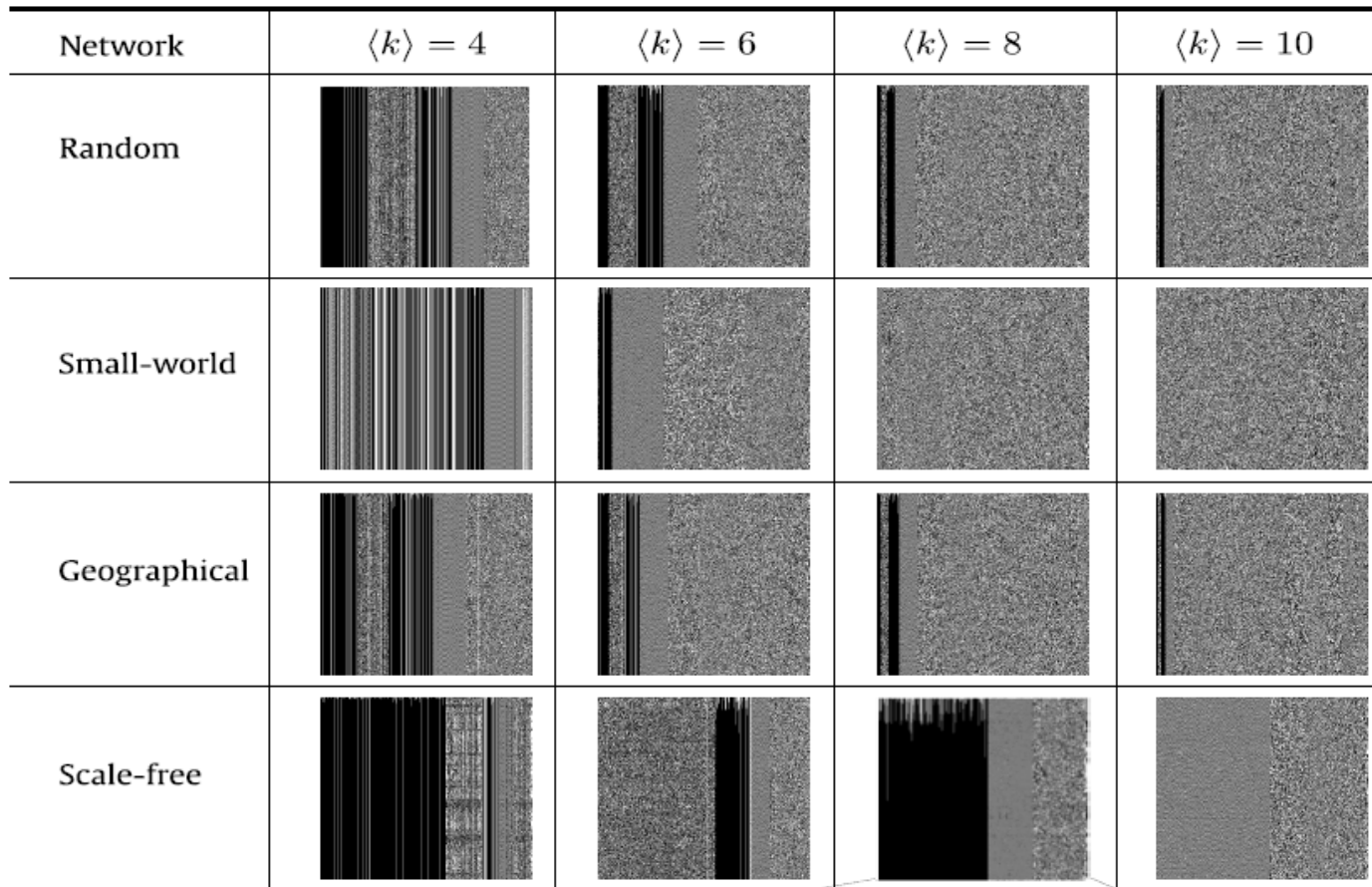
$s(c_i, 2) = 1$ and $22\% \not\leq \rho(c_i) < 33\%$
or
 $33\% \not\leq \rho(c_i) < 44\%$

⇓ die



$s(c_i, 3) = 0$

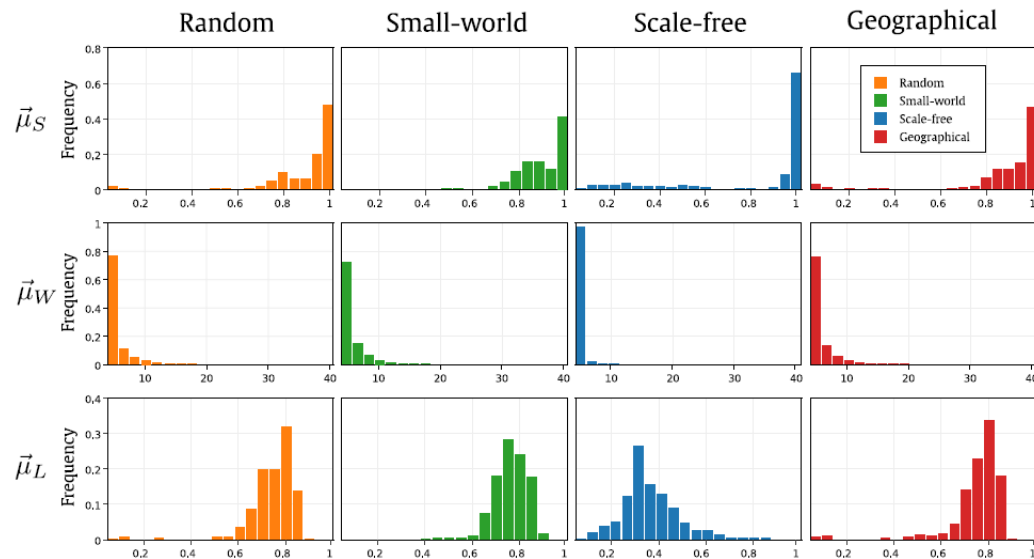
Spatio-temporal patterns



Spatial-time diagram of the life-like network-automaton of the 4 synthetic network models with $N = 500$ vertices and different degree, evolved by $t = 500$ iterations using rule B1357 / S2468

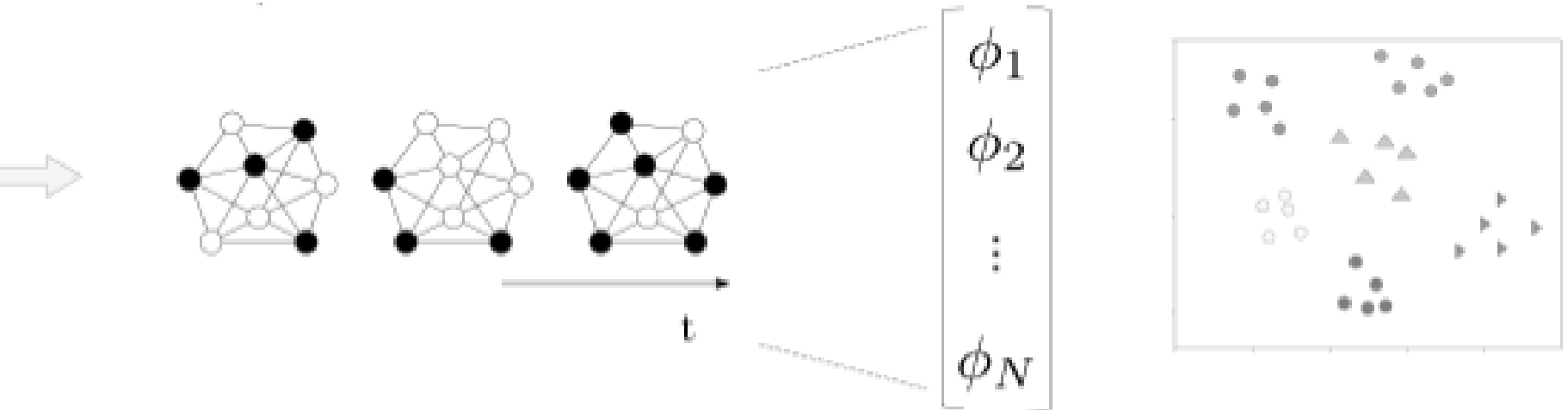
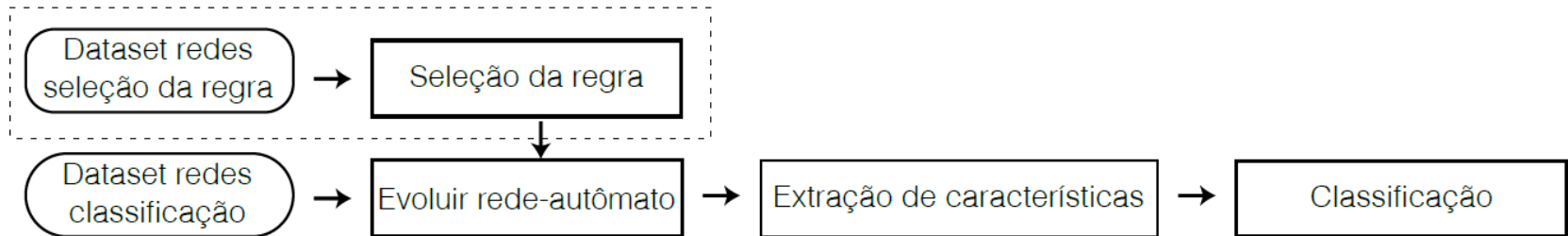
LLNA's characterization

- Shannon entropy distribution $\vec{\mu}_S$ $H_{S_i} = -(p_i^0 \log_2 p_i^0 + p_i^1 \log_2 p_i^1)$
- Word length distribution $\vec{\mu}_W$ $H_{W_i} = -\sum_{l=1}^L p_i^l \log_2 p_i^l$
- Lempel Ziv complexity distribution $\vec{\mu}_L$



Histogram of the three distributions used to quantitatively analyze the spatio-temporal patterns of distinct network models: Shannon entropy $\vec{\mu}_S$, word length $\vec{\mu}_W$ and Lempel-Ziv complexity $\vec{\mu}_L$. The following parameters were adopted: $N = 500$, $\langle k \rangle = 4$ and $t = 350$.

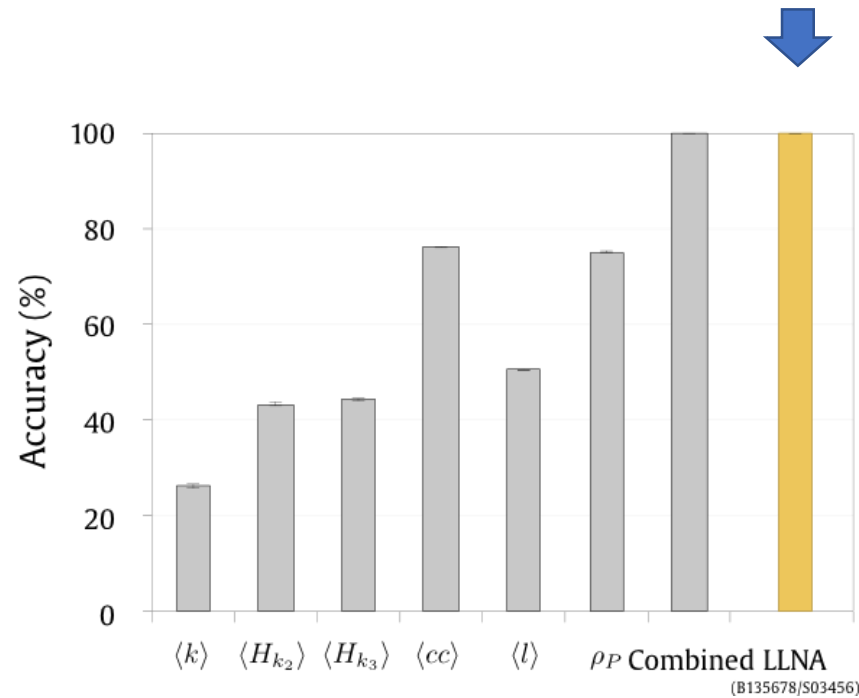
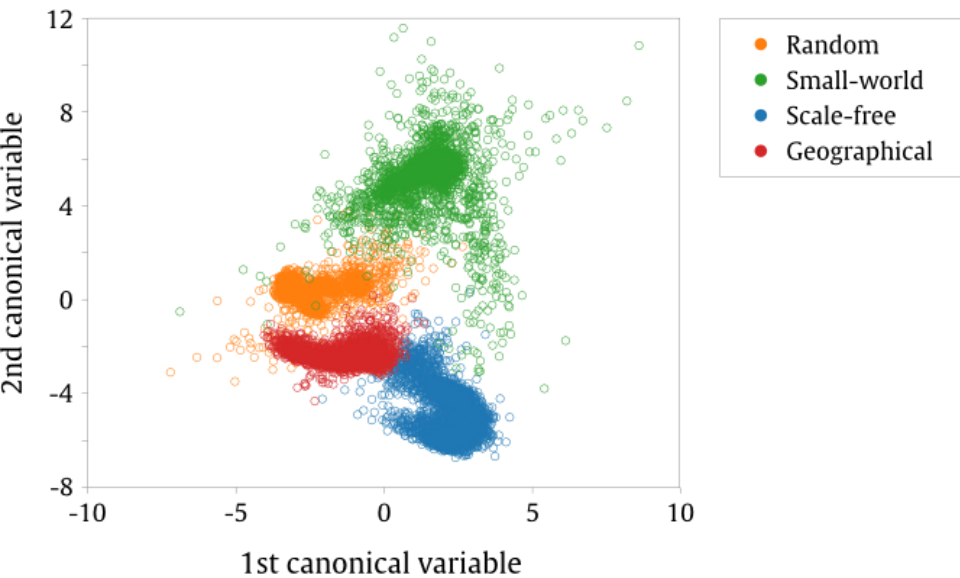
Pattern recognition based on LLNA



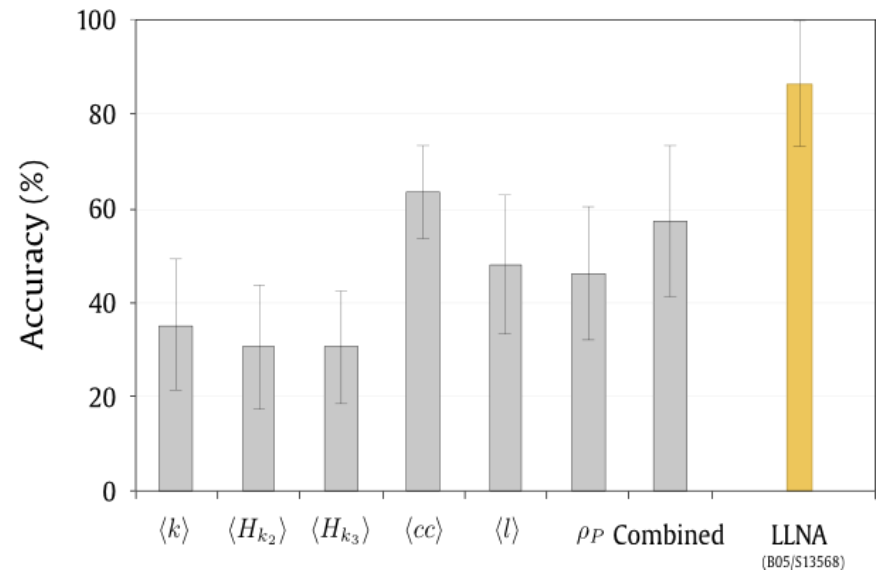
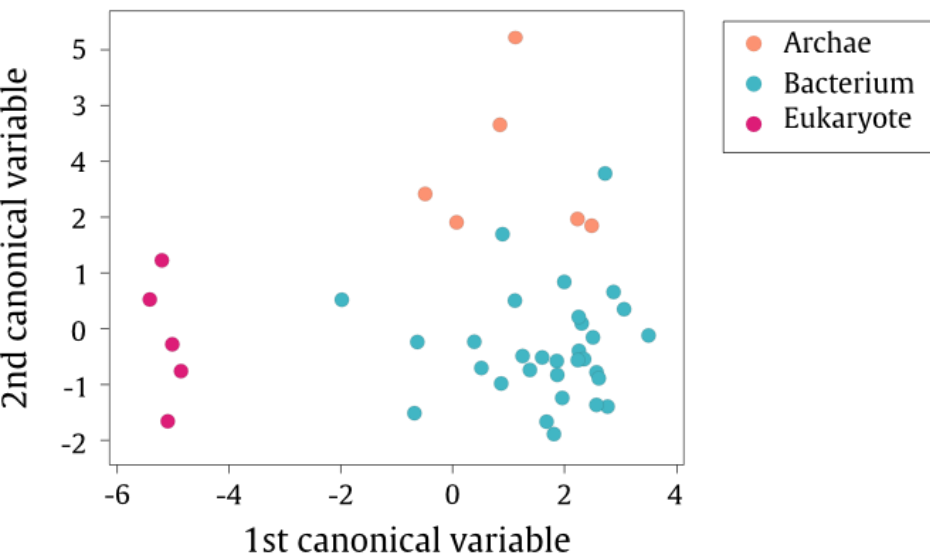
Datasets

	#Classes / Legendas	<u>Validação</u>	<u>Seleção da regra</u>
		#Redes × classe	#Redes × classe
Redes sintéticas			
Sintética	4 ER, WS, BA, GEO	11200 2800	1400 350
Redes metabólicas			
Jeong	3 Archae, Bacterium, Eukarvotte	37 4, 30 e 3	6 2, 2 e 2
Redes sociais			
SNAP	2 Twiiter, Google+	100 50	30 15
Redes textuais			
Autoría	8 Doyle, Stoker, Darwin, Dickens, Hardy, Wo-dehouse, Poe, Munro	40 5	60 5
Redes de estômato			
<i>Trandescantia</i>	3 natural, L4h, L24h	12 4	6 2

Classification of synthetic network models

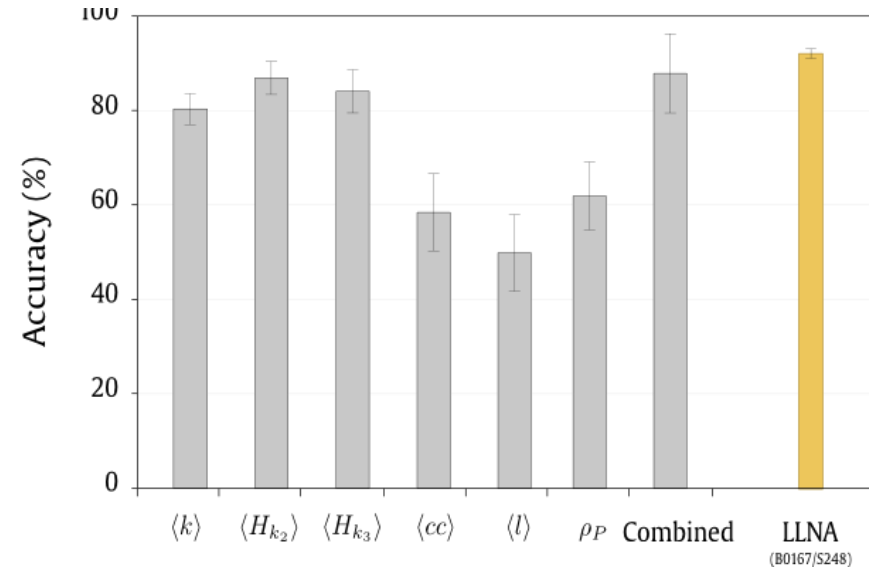
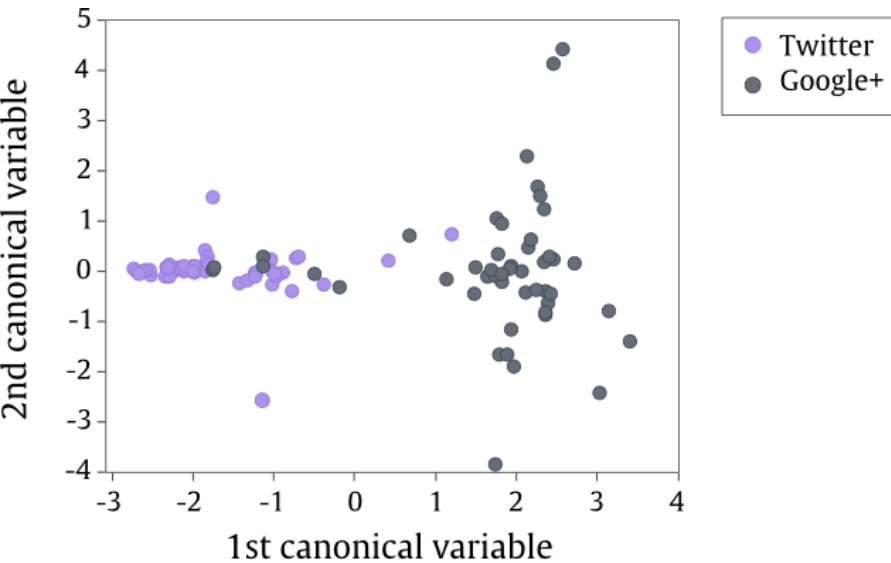


Identification of organisms from distinct domains of life



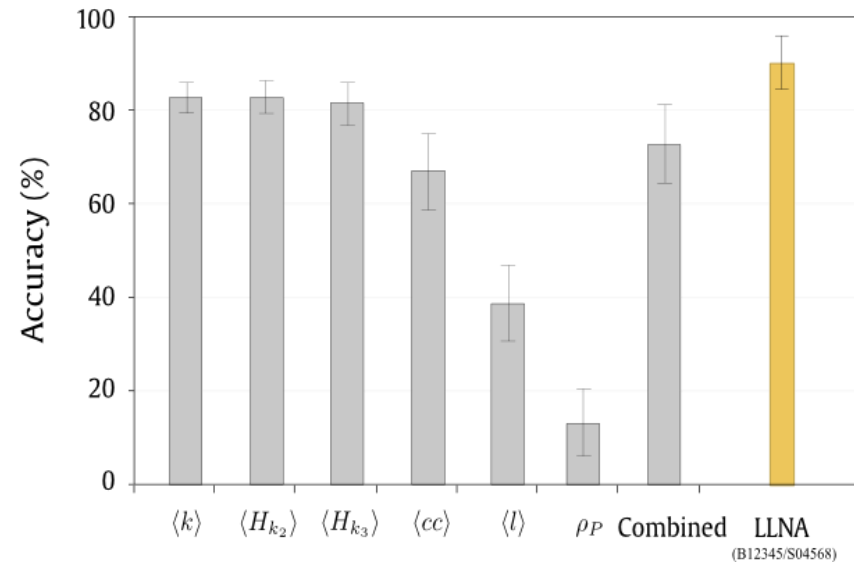
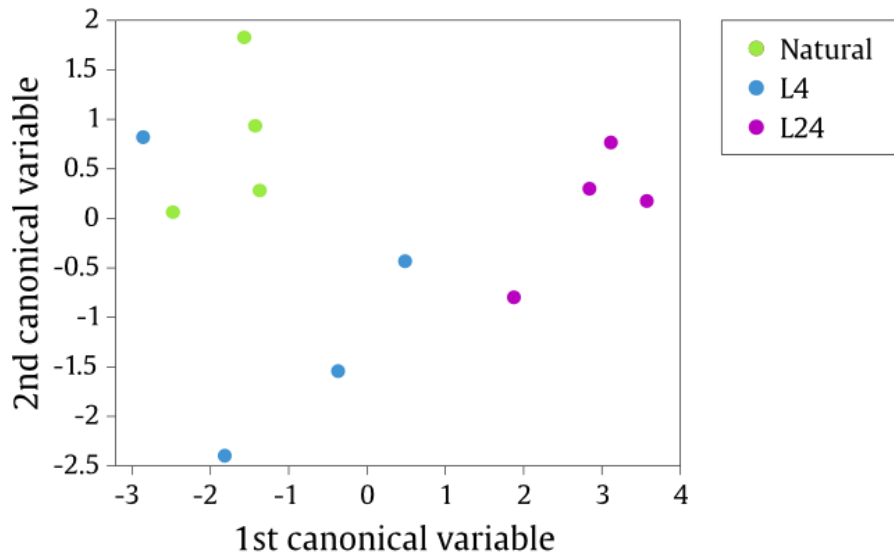
Dataset Jeong et. al. [18]

Identification of structural patterns in social networks

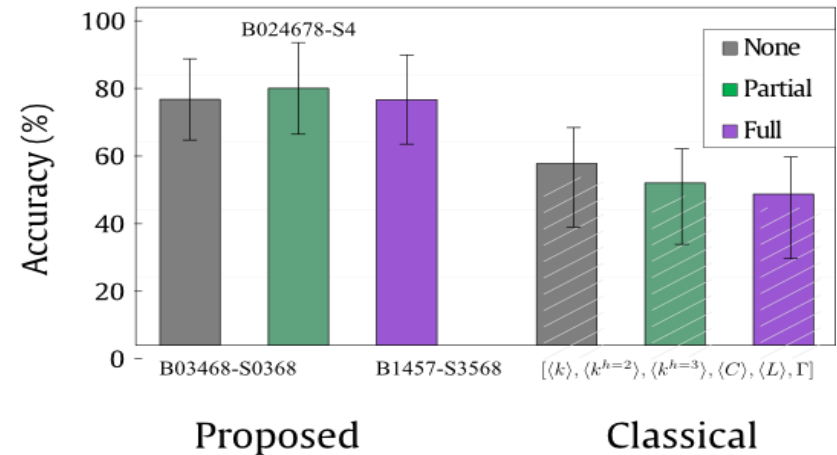
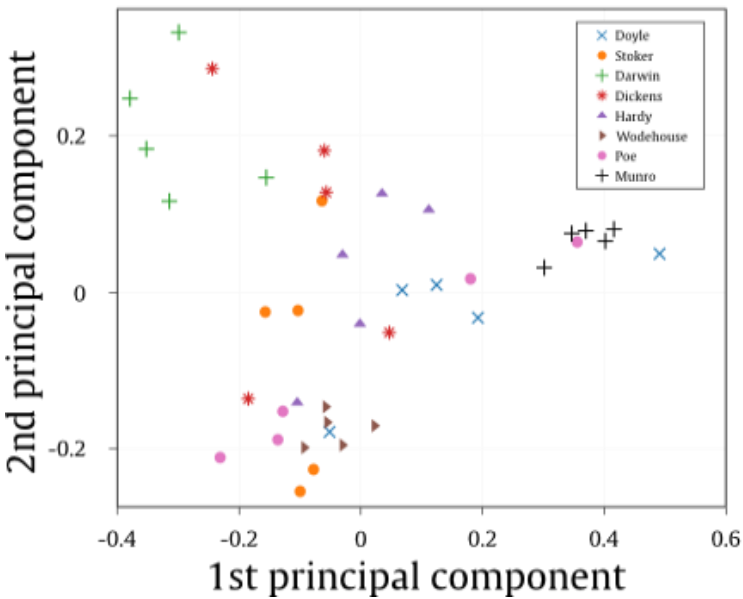


Dataset SNAP [53]

Classifying stomata distribution patterns varying according to lighting conditions



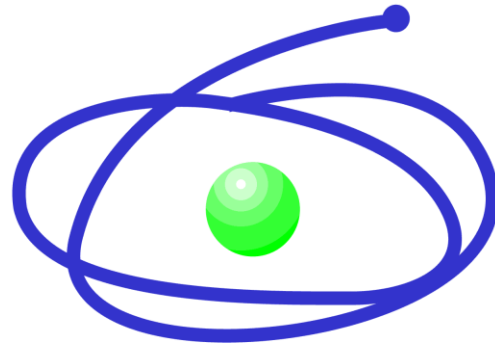
Authorship attribution



		Predicted Class							
		Doyle	Stoker	Darwin	Dickens	Hardy	Wodeh.	Poe	Munro
Real Class	Doyle	2			1	1			1
	Stoker		4					1	
	Darwin			5					
	Dickens	1			4				
	Hardy	1	1			3			
	Wodeh.						5		
	Poe							5	
	Munro								5

References

- [1] ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, v. 74, n. 1, p. 47–97, 2002.
- [3] BAETENS, J. M.; DE BAETS, B. Phenomenological study of irregular cellular automata based on Lyapunov exponents and Jacobians. *Chaos*, v. 20, n. 3, p. 033112, 2010.
- [7] COSTA, L. D. F.; RODRIGUES, F. A.; TRAVIESO G.; BOAS, P. R. V. Characterization of complex networks: a survey of measurements. *Advances In Physics*, v. 56, n. 1, p. 167, 2007.
- [10] W.N. Gonc,alves, B.B. Machado, and O.M. Bruno. A complex network approach for dynamic texture recognition. *Neurocomputing*, 153(0):211–220, 2015.
- [11] DOGANAKSOY, A.; GÖLOGLU, F. On Lempel-Ziv complexity of sequences. In: GONG, G. et.al., (Ed.) *Sequences and their applications – SETA 2006*. Berlin, Heidelberg: Springer, 2006 p. 180–189. (Lecture notes in computer science, v. 4086).
- [13] ERDÖS, P.; RÉNYI, A. On random graphs I. *Publicationes Mathematicae Debrecen*, v. 6, p. 290–297, 1959. Disponível em: <http://ftp.math-inst.hu/~p_erdos/1959-11.pdf>. Acesso em: 20.10.2015.
- [18] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A.-L. The large-scale organization of metabolic networks. *Nature* 407, 651–654 (2000).
- [21] Miranda GHB, Machicao J, Bruno OM. Exploring Spatio-temporal Dynamics of Cellular Automata for Pattern Recognition in Networks. *Scientific Reports*. 2016;6(37329).
- [28] WATTS, D. J.; STROGATZ, S. H. Collective dynamics of “small-world” networks. *Nature*, v. 393, n. 6684, p. 440–442, 1998.
- [53] Leskovec, J. & Krevl, A. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data> (2014).



C A P E S