A non-linear approach for pattern recognition in networks

Marina Jeaneth Machicao Justo

Advisor: Prof. Dr. Odemir M. Bruno

18 de setembro de 2017

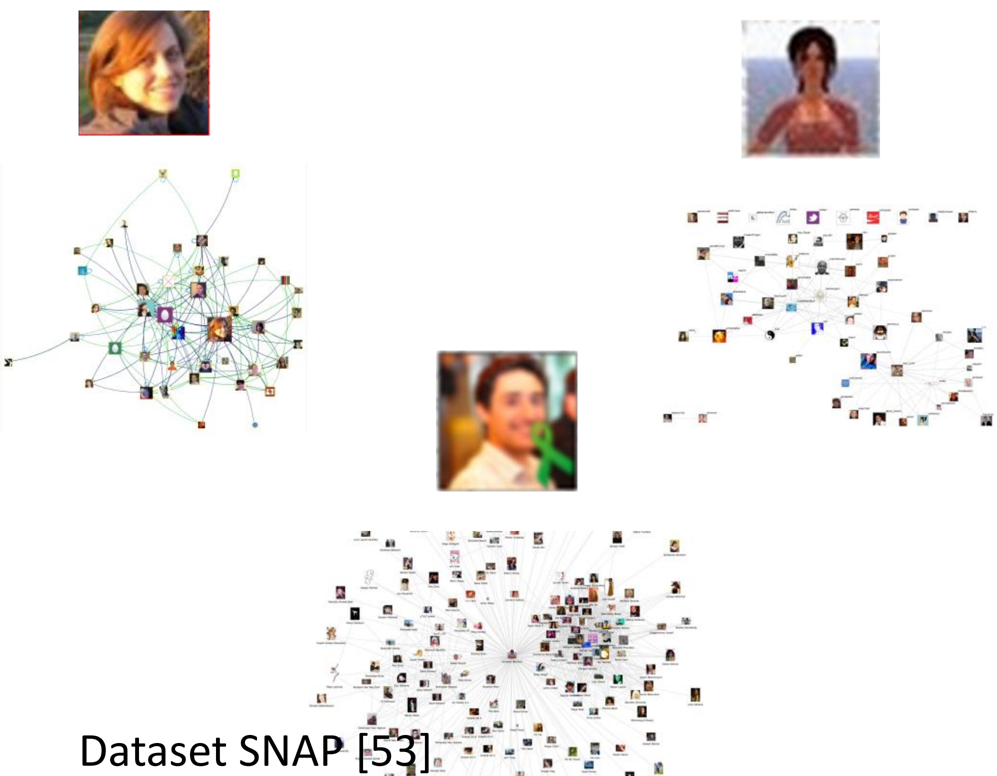
Ph. D Project Profile

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

Networks are everywhere!

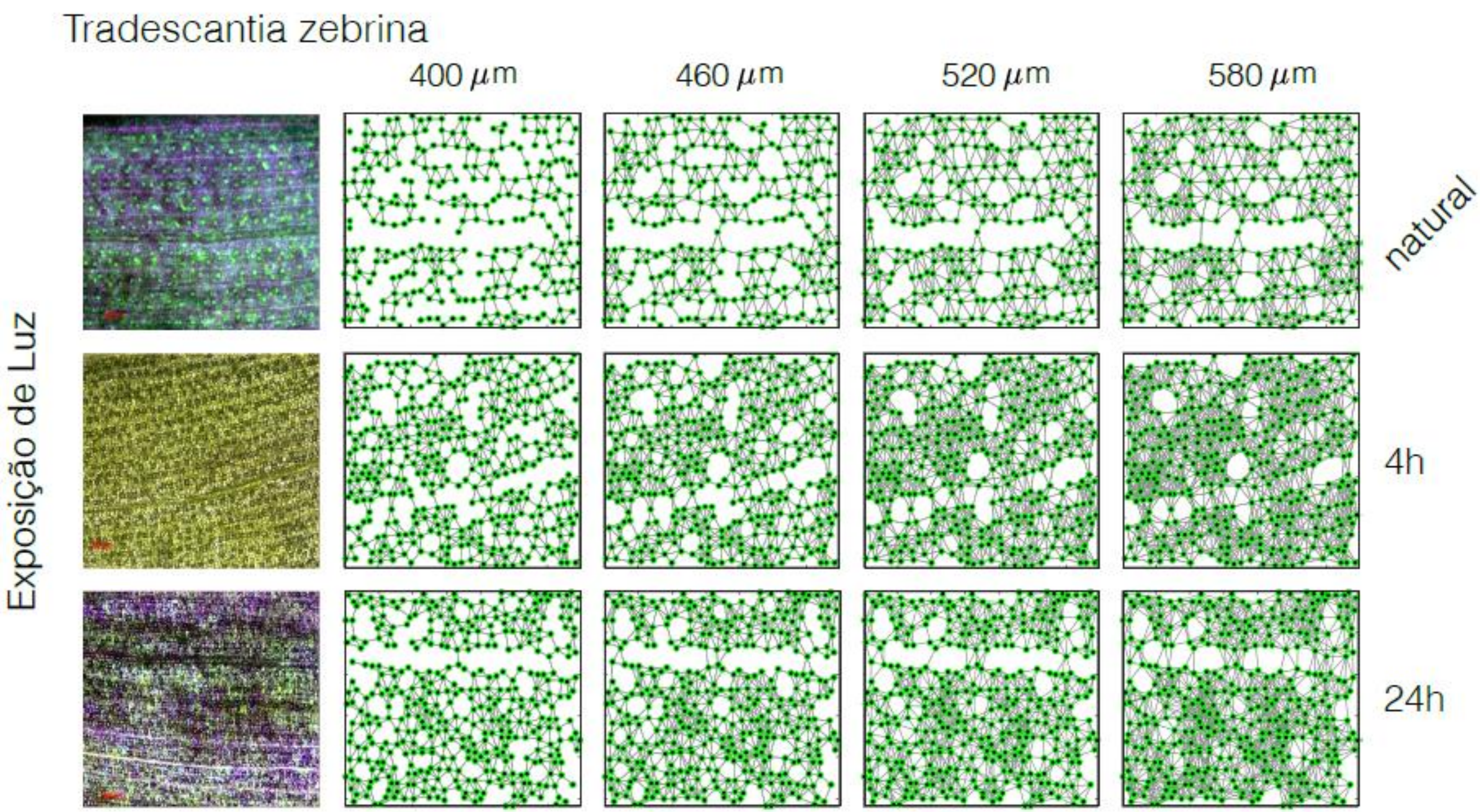
Social networks

- Twitter
- Google +



Dataset SNAP [53]

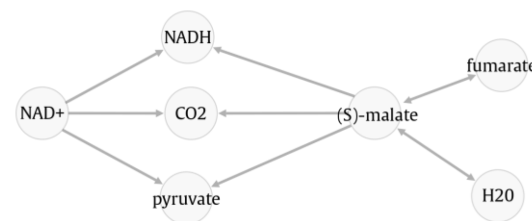
Stomata distribution networks



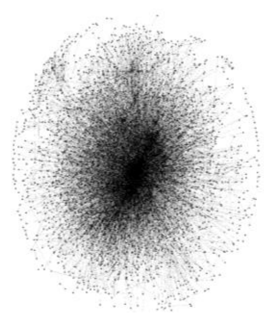
Metabolic networks

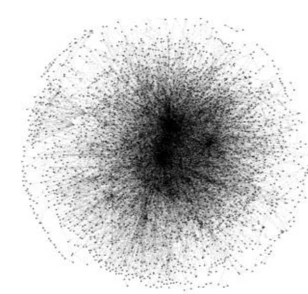
- Metabolites (nodes) relationship are connected according to the reaction directions.

(S)-malate + NAD+ ⟶ CO2 + pyruvate + NADH
(S)-malate ⟷ fumarate + H20

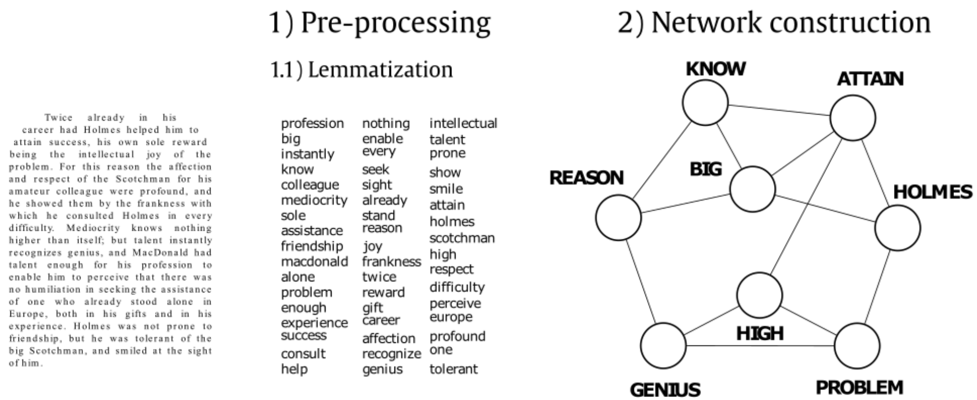


- Organisms [18]:
  - Archaea
  - Bacterium
  - Eukaryote

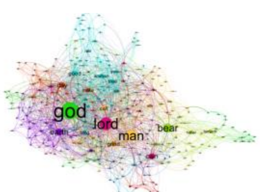


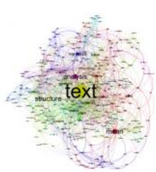


Textual networks

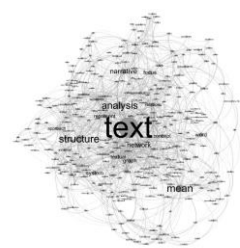


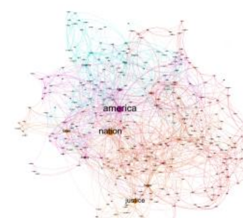
Poe





Darwin





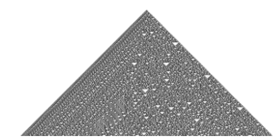
Pattern recognition in networks

- 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)

Specie *conus textile*

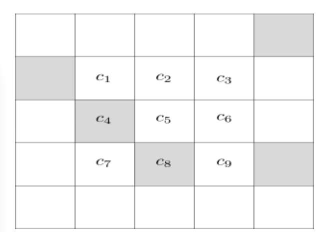


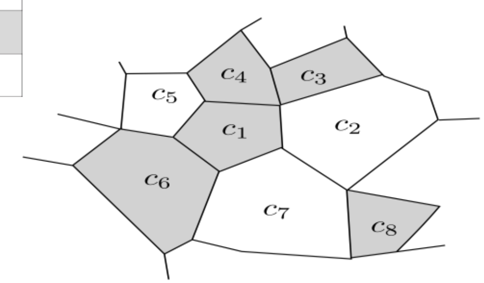
Rule 30 ECA

- Represented by

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

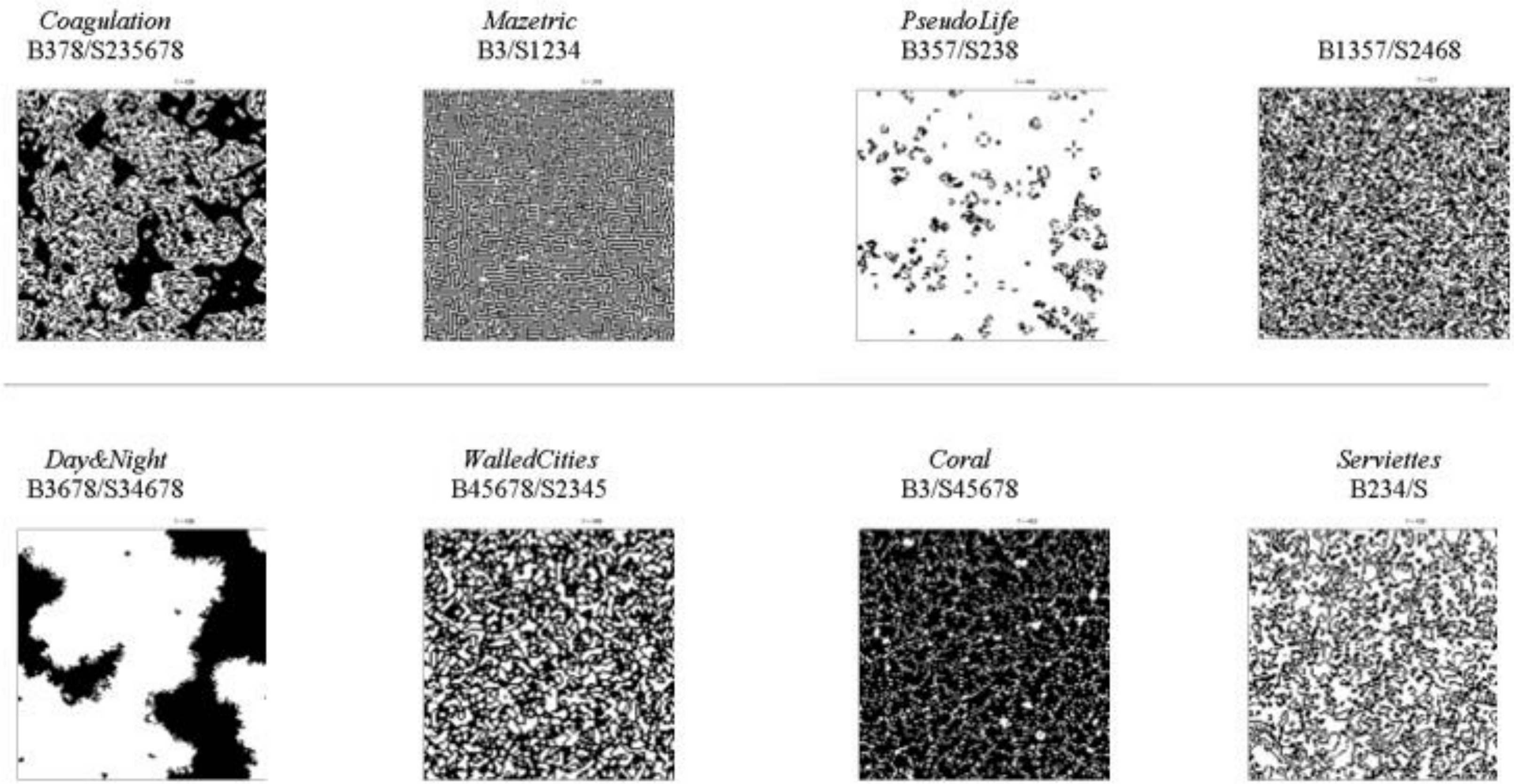
- $\mathcal{T}$: tesselação
- S : Estados, $S = \{s_0, s_1, \ldots, s_{k-1}\}$
- s : função de estado, $s(c_i, t)$
- $\mathcal{N}$: função vizinhança
- ϕ : regra de transição





Life-like CA

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



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

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

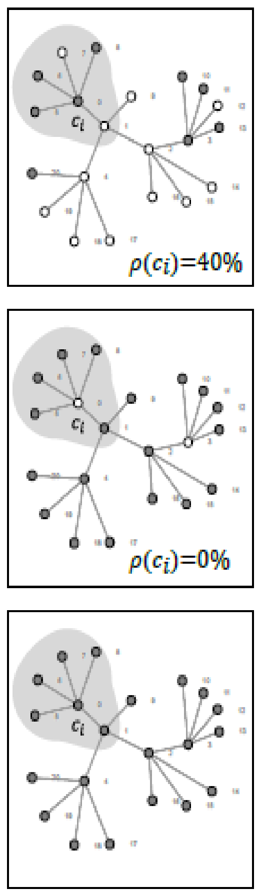
Life-like network automata (LLNA)

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

t=1

t=2

t=3



Rule: B3/S23

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

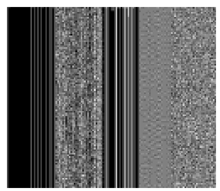
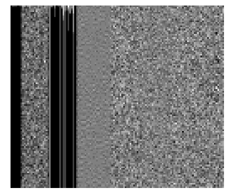
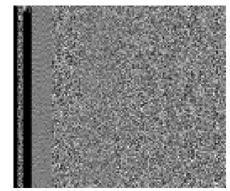
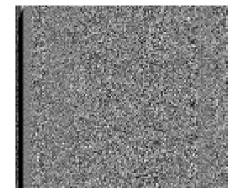
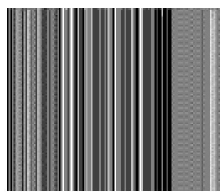
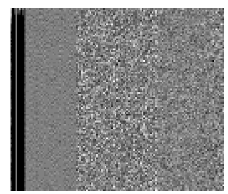
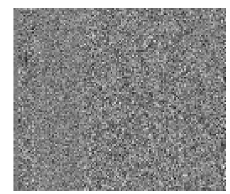
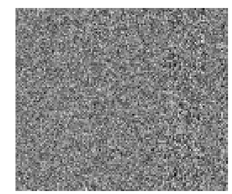
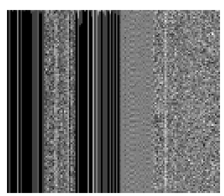
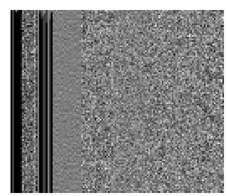
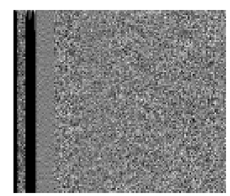
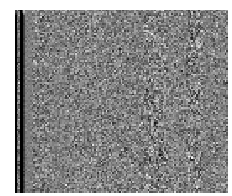
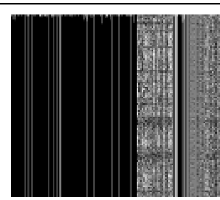
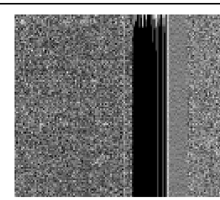
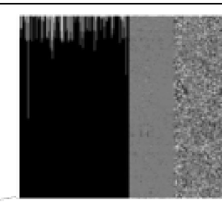
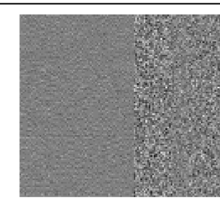
$\Downarrow$ born

$s(c_i, 2) = 1$ and $22\% \nleq \rho(c_i) \nless 33\%$ or $33\% \nleq \rho(c_i) \nless 44\%$

$\Downarrow$ die

$s(c_i, 3) = 0$

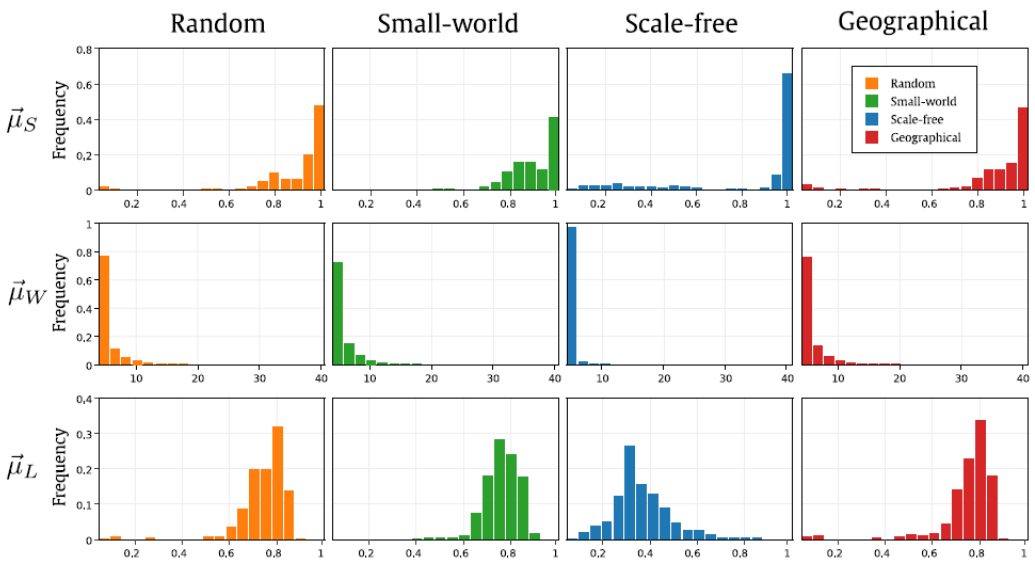
Spatio-temporal patterns

Network	$\langle k \rangle = 4$	$\langle k \rangle = 6$	$\langle k \rangle = 8$	$\langle k \rangle = 10$
Random				
Small-world				
Geographical				
Scale-free				

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

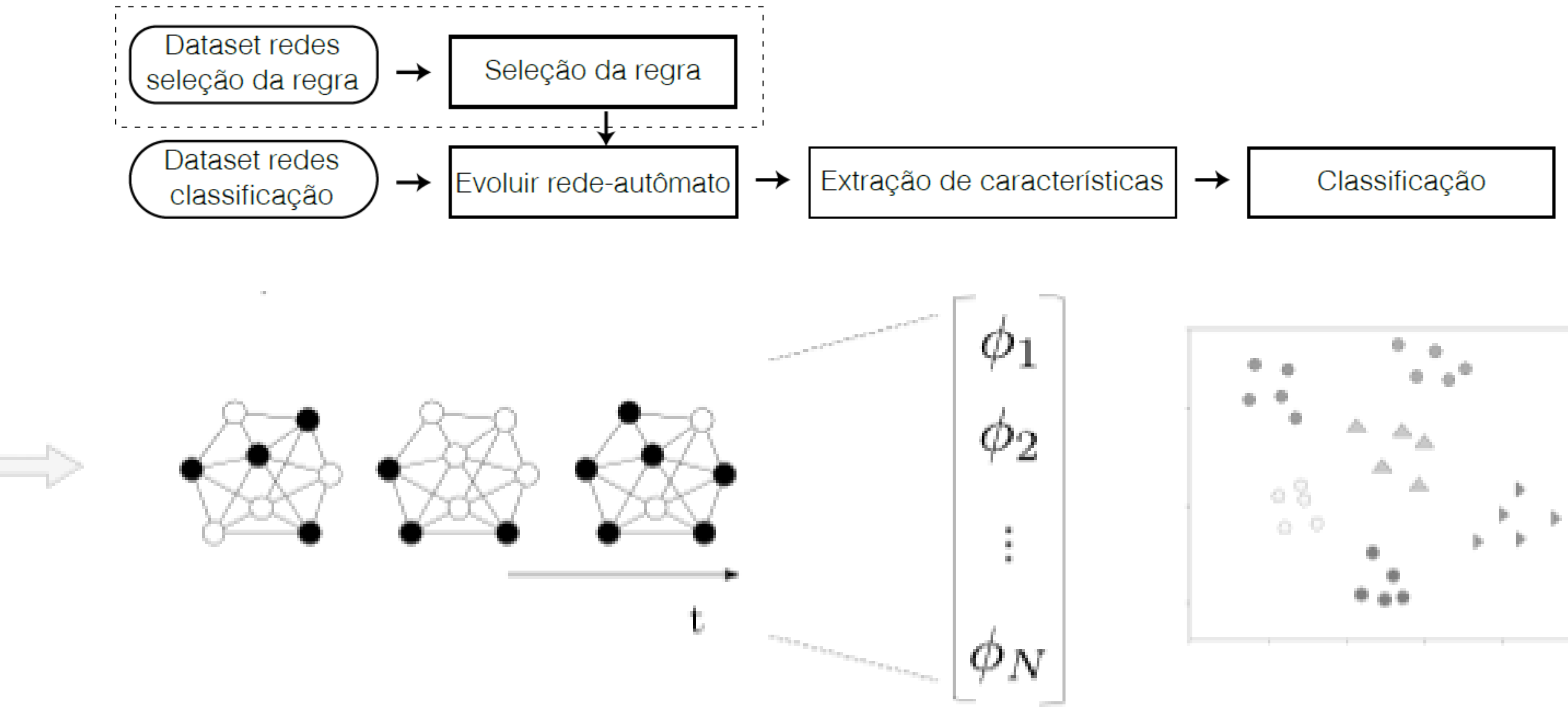
LLNA's characterization

- 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 $t = 350$.

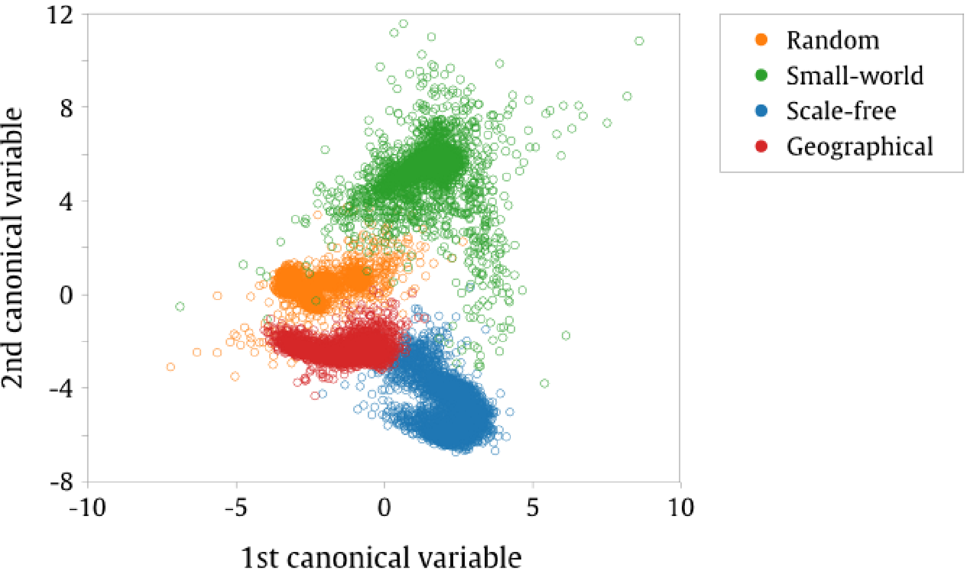
Pattern recognition based on LLNA

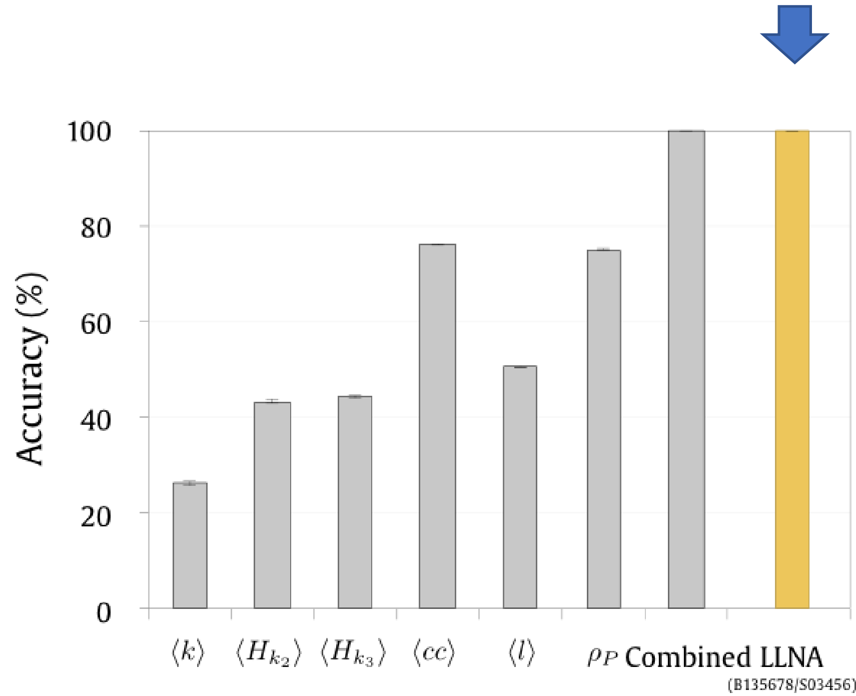


Datasets

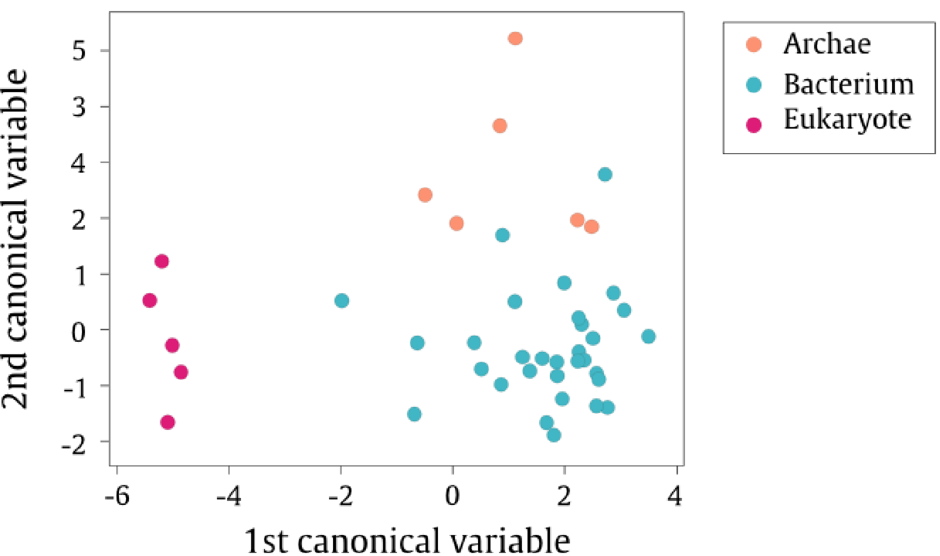
	#Classes / Legendas		Validação #Redes	× classe	Seleção da regra #Redes	× classe
Redes sintéticas						
Sintética	4	ER, WS, BA, GEO	11200	2800	1400	350
Redes metabólicas						
Jeong	3	Archae, Bacterium, Eukaryote	37	4, 30 e 3	6	2, 2 e 2
Redes sociais						
SNAP	2	Twiiter, Google+	100	50	30	15
Redes textuais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

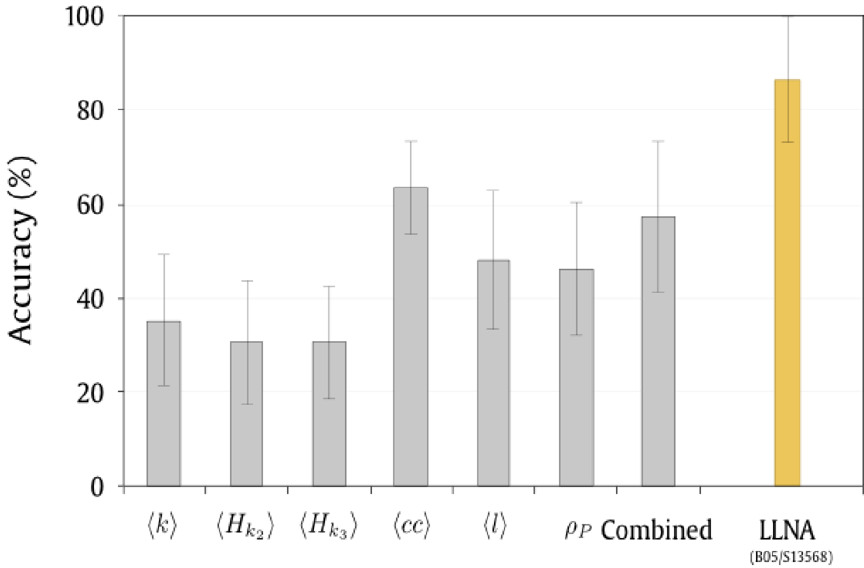
Classification of synthetic network models





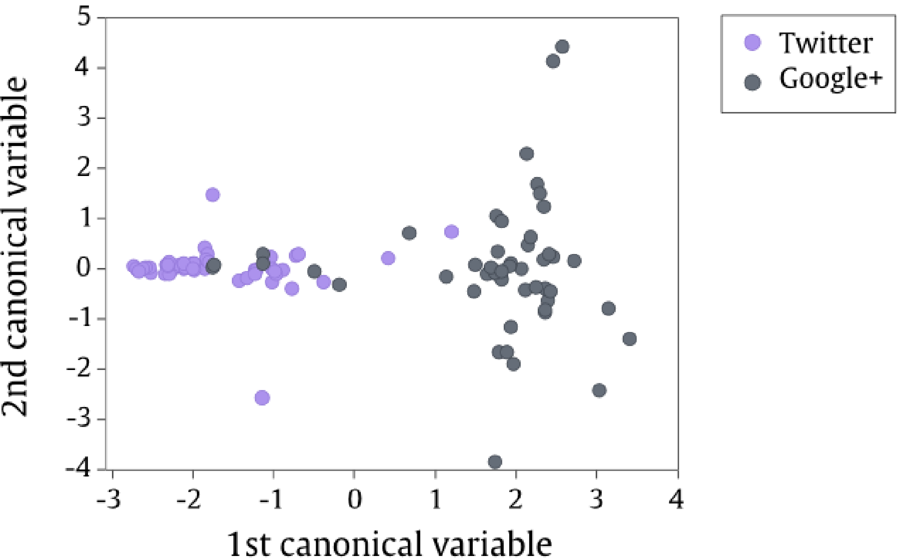
Identification of organisms from distinct domains of life

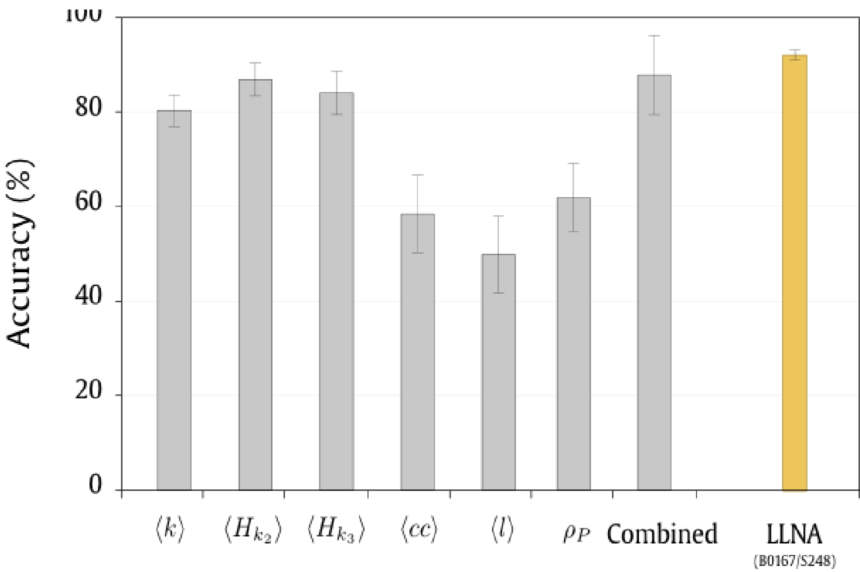




Dataset Jeong et. al. [18]

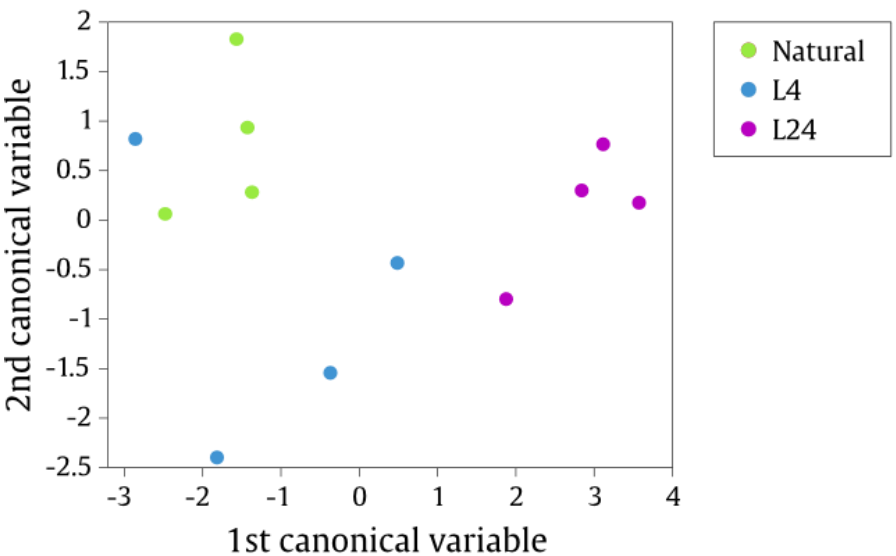
Identification of structural patterns in social networks

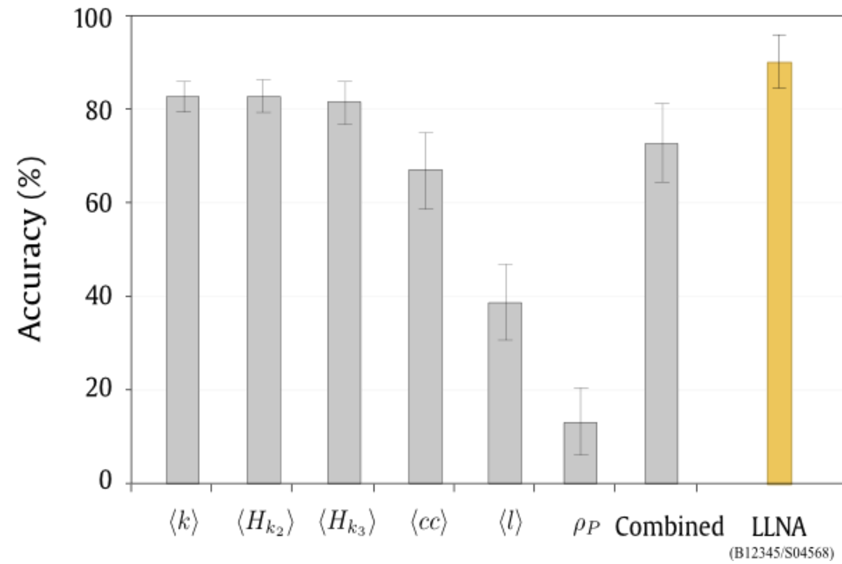




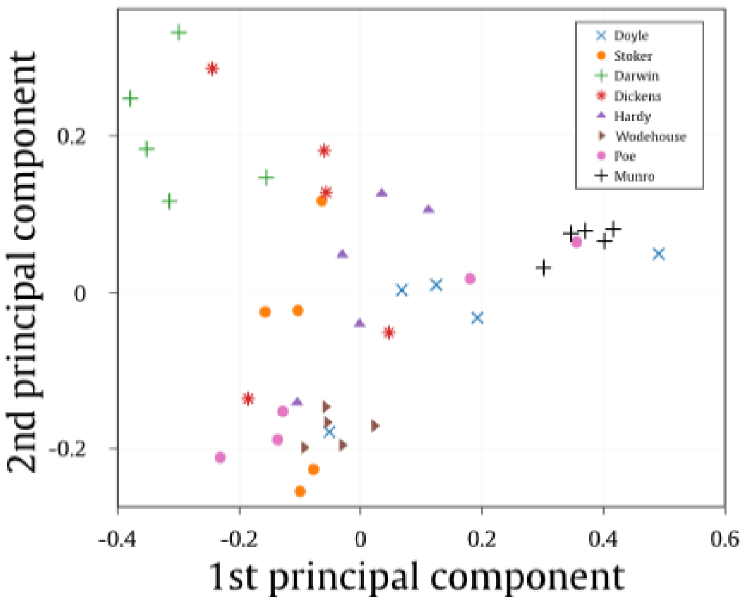
Dataset SNAP [53]

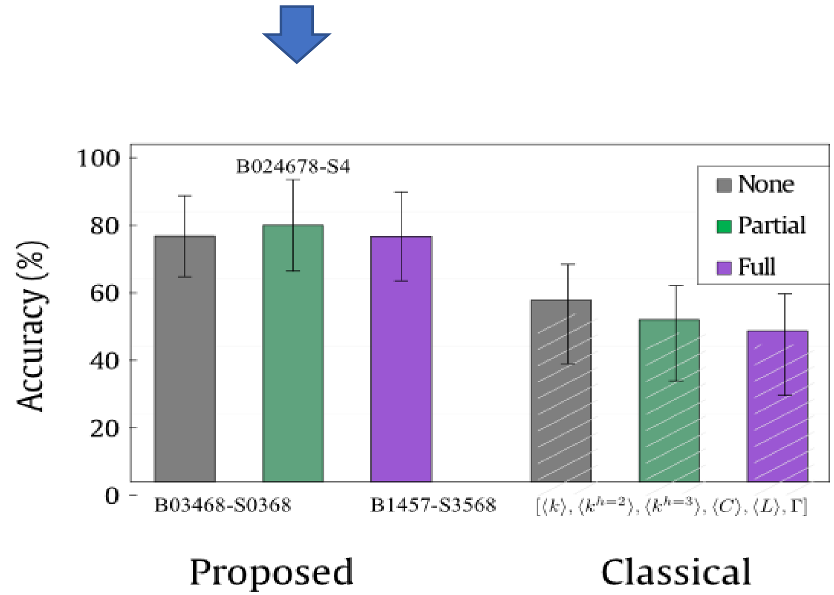
Classifying stomata distribution patterns varying according to lighting conditions





Authorship attribution





Predicted Class

Real Class	Doyle	Stoker	Darwin	Dickens	Hardy	Wodeh.	Poe	Munro
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