

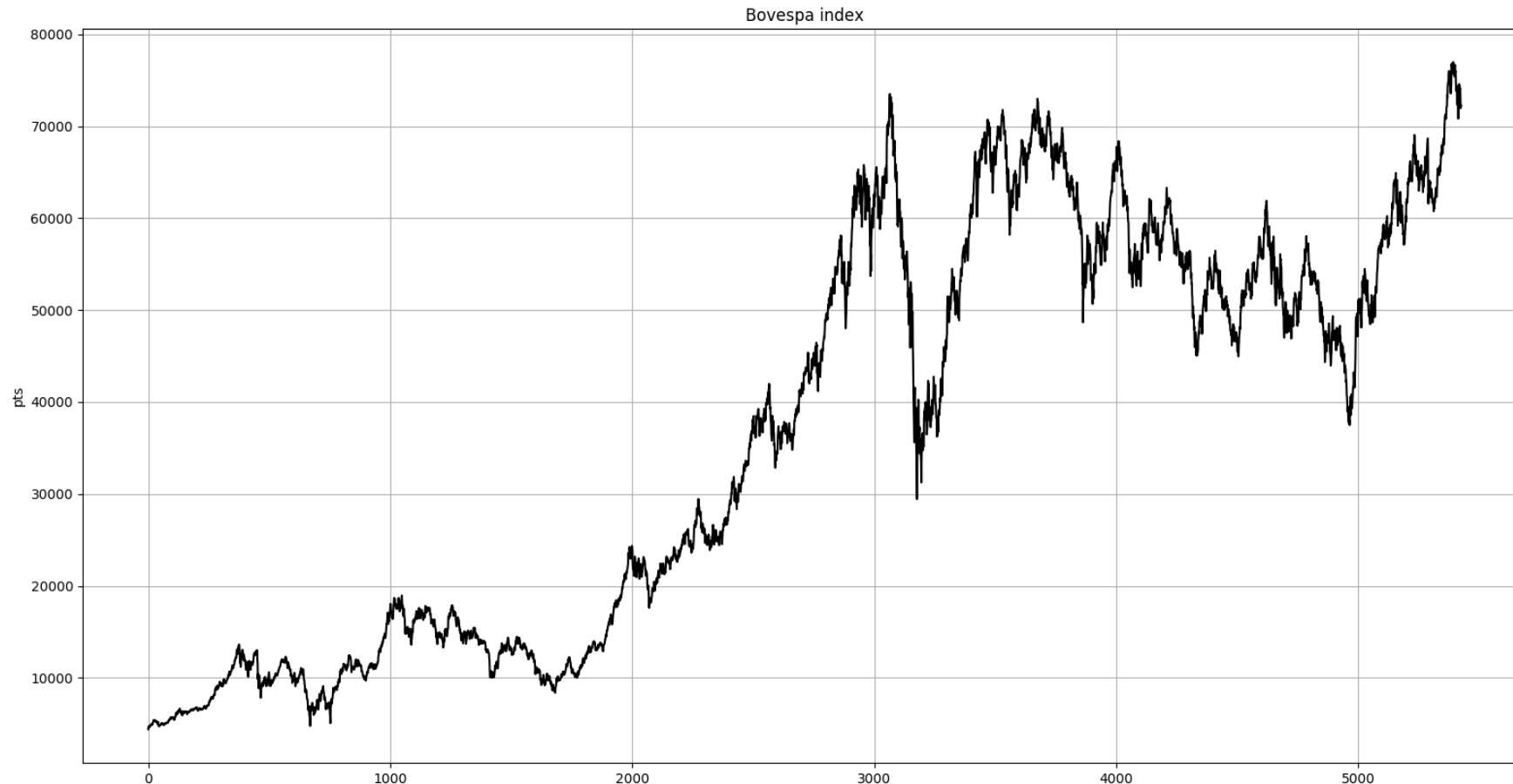
Time Series Trend Detection and Forecasting Using Complex Network Topology Analysis

- Leandro Anghinoni – Universidade de São Paulo – DCM/RP
- Liang Zhao – Universidade de São Paulo – DCM/RP

AGENDA

- Introdução
- Método proposto
 - Classificação
 - Previsão
- Resultados
 - Curva artificial
 - Curva real
- Próximos passos

- ❑ Como classificar e prever a tendência de séries financeiras?



Este problema pode ser abordado de duas formas:

1. Previsão de valores:

- Prevê um valor dentro de uma margem de confiança
- Normalmente feita com algum modelo de regressão
 - ARIMA
 - ARIMA + SVM
 - ARIMA + NN

2. Previsão de tendência:

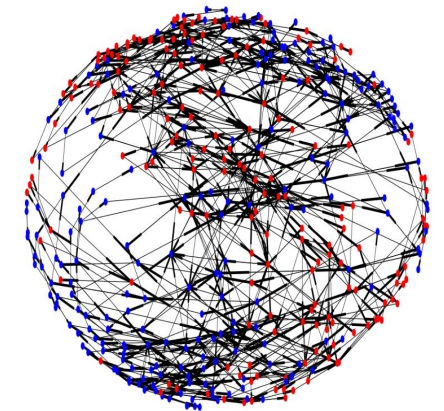
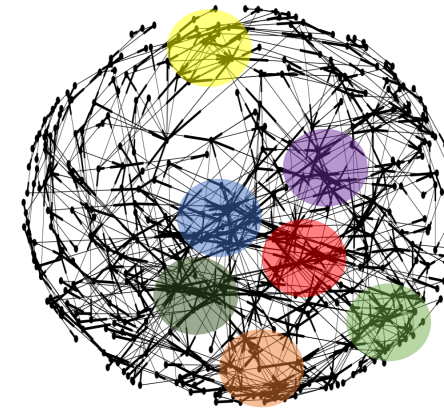
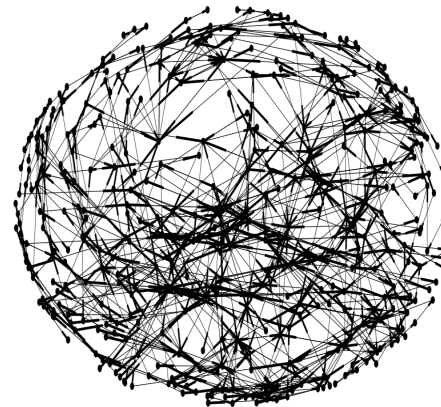
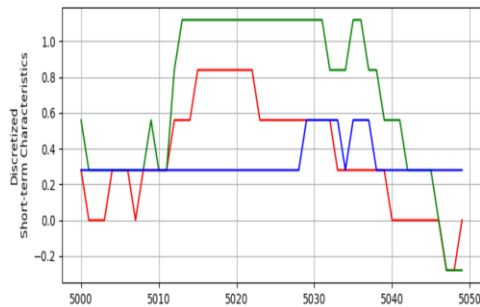
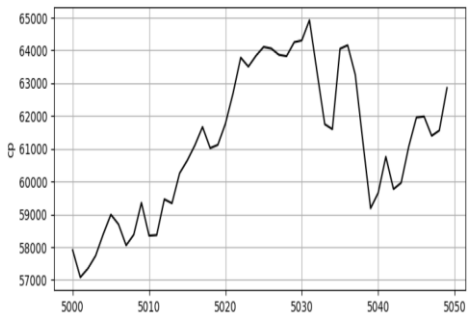
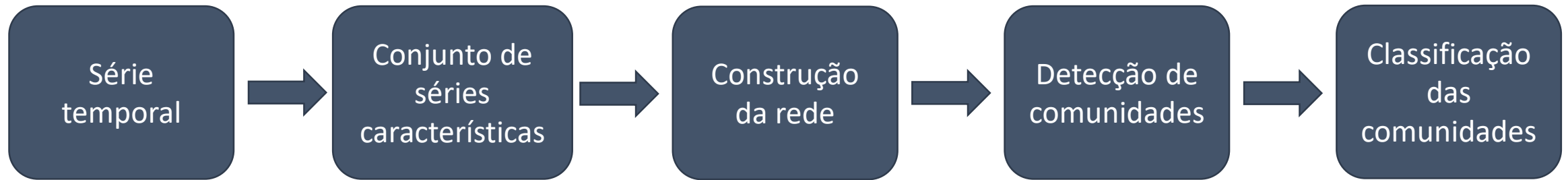
- Prevê uma tendência (normalmente alta, baixa, lateral)
- Pode ser feita com classificadores (DT, RF, MLP, RNN, LSTM, etc...)

MAS A BASE DE DADOS PRECISA ESTAR ROTULADA → O QUE É MUITO SUBJETIVO

- ❑ PROPOSTA:
- ❑ Framework para classificação baseado em clusterização não supervisionada
 - ❑ Algumas vantagens...
 - ❑ A base não precisa ser rotulada manualmente ou com base em um indicador
 - ❑ Tendências tem duração variável e não se limitam a 3
 - ❑ Tendência pode ser medida em relação ao seu retorno histórico
 - ❑ Capacidade de reconhecer padrões do algoritmo
 - ❑ Probabilidade de mudança de tendência pode ser observada na rede
 - ❑ Topologia da rede gera dados para novas formas de previsão

MÉTODO PROPOSTO

□ Framework de classificação



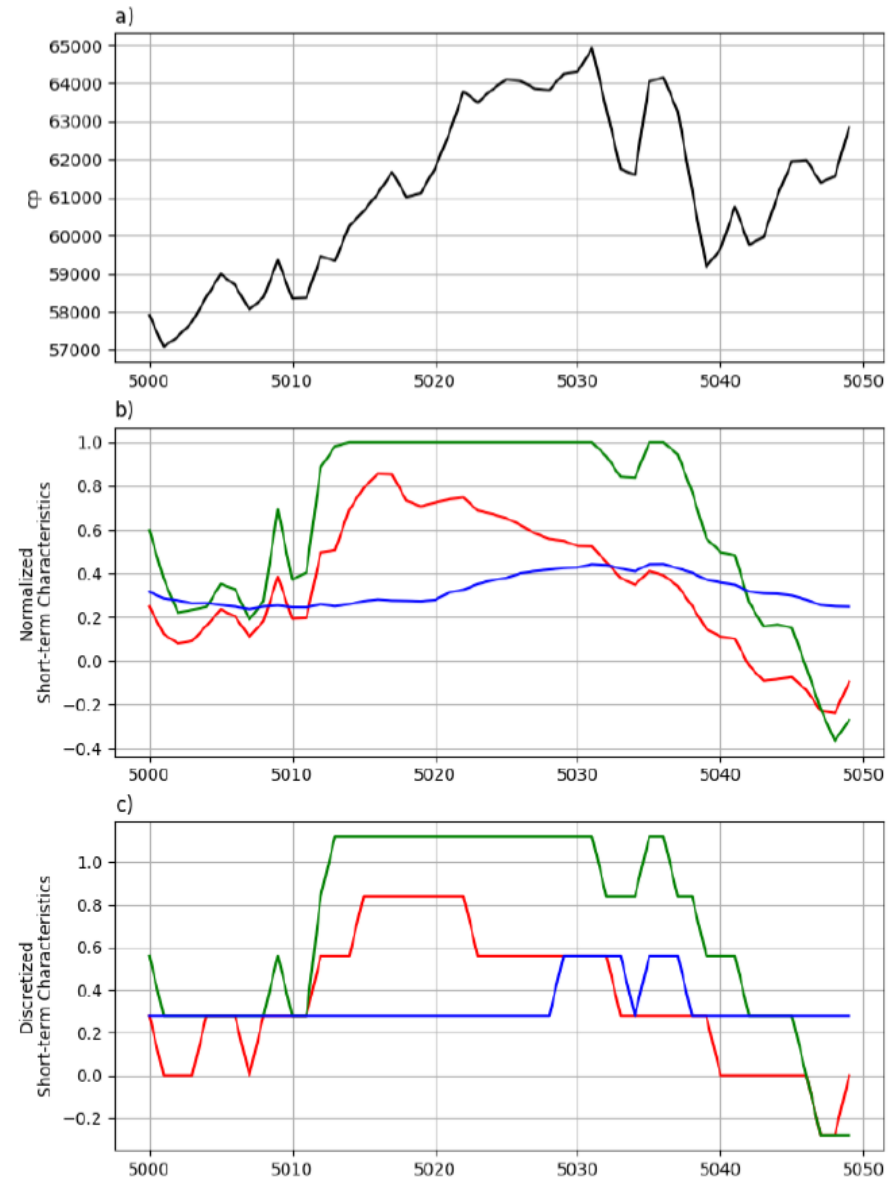
MÉTODO PROPOSTO

- ❑ Extração de características
- ❑ 6 características foram usadas:

$$noise_t = \frac{cpt_t - MA(q)_t}{MA(q)_t}$$

$$grad_t = \frac{MA(q)_t - MA(q)_{t-1}}{MA(q)_{t-1}}$$

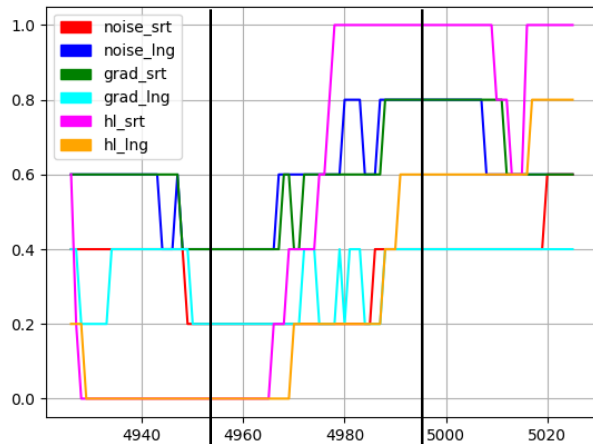
$$rhl_t = \frac{cpt_t - \text{MIN}(cpt_{t-q}:cpt_t)}{\text{MAX}(cpt_{t-q}:cpt_t) - \text{MIN}(cpt_{t-q}:cpt_t)}$$



MÉTODO PROPOSTO

❑ Criação da rede (com peso nas arestas e direcionada)

❑ Método: redes de transição (entre estados do processo) [1]



→ $Node_B = [0.4, 0.8, 0.8, 0.4, 1.0, 0.6]$

→ $Node_A = [0.2, 0.4, 0.4, 0.2, 0.0, 0.0]$

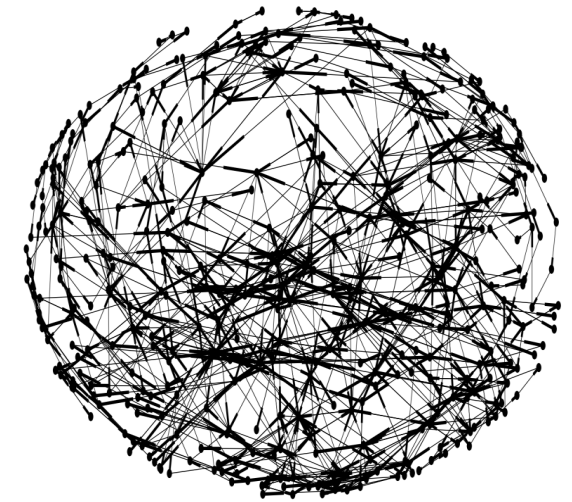
❑ $G(n,l)$

❑ $nodes$: estados do processo

❑ $node$: vetor de estado no tempo t

❑ $node \rightarrow [c1, c2, c3, c4, c5, c6]_t$

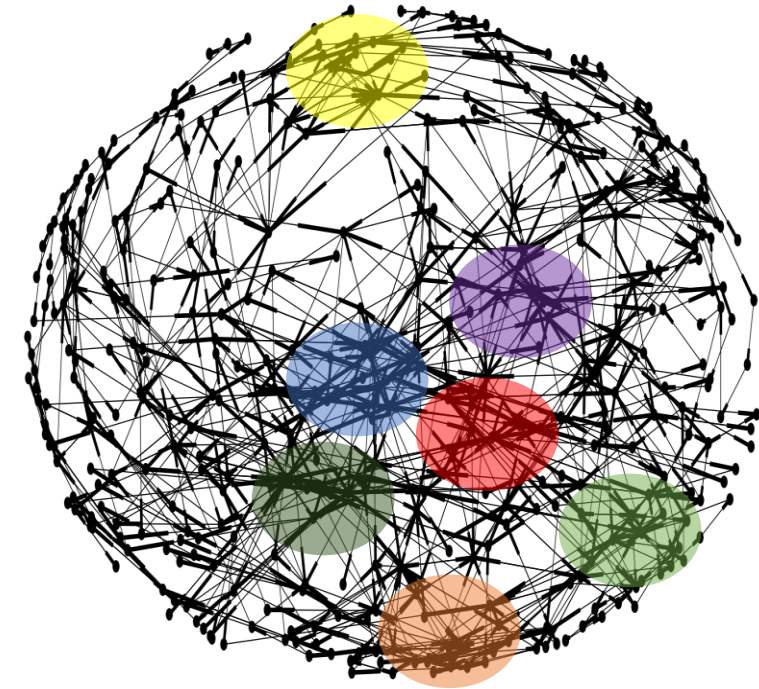
❑ $link$: aresta $node_t \rightarrow node_{t+1}$



[1] Donner, R.V., Small, M., Donges, J.F., Marwan, N., Zou, Y., Xiang, R. and Kurths, J., 2011. Recurrence-based time series analysis by means of complex network methods. International Journal of Bifurcation and Chaos, 21(04), pp.1019-1046.

MÉTODO PROPOSTO

- ❑ Detecção de comunidades
 - ❑ Método aglomerativo baseado na modularidade (Q) [2]
 - ❑ Algoritmo para quando Q para de aumentar
 - ❑ Q representa a diferença de estrutura para uma rede aleatória
 - ❑ Redes com $Q > 0.3$ podem ser consideradas estruturadas, ou seja, os links não são aleatórios

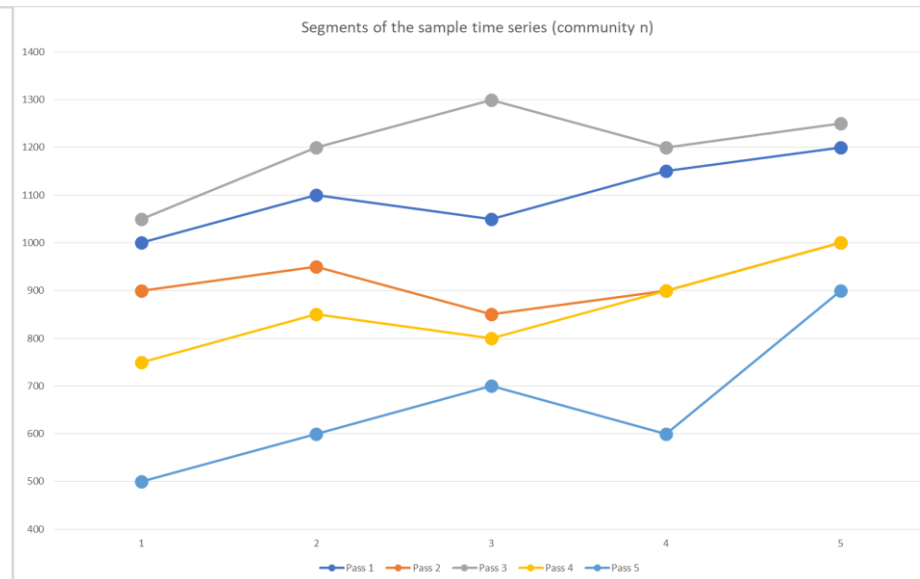
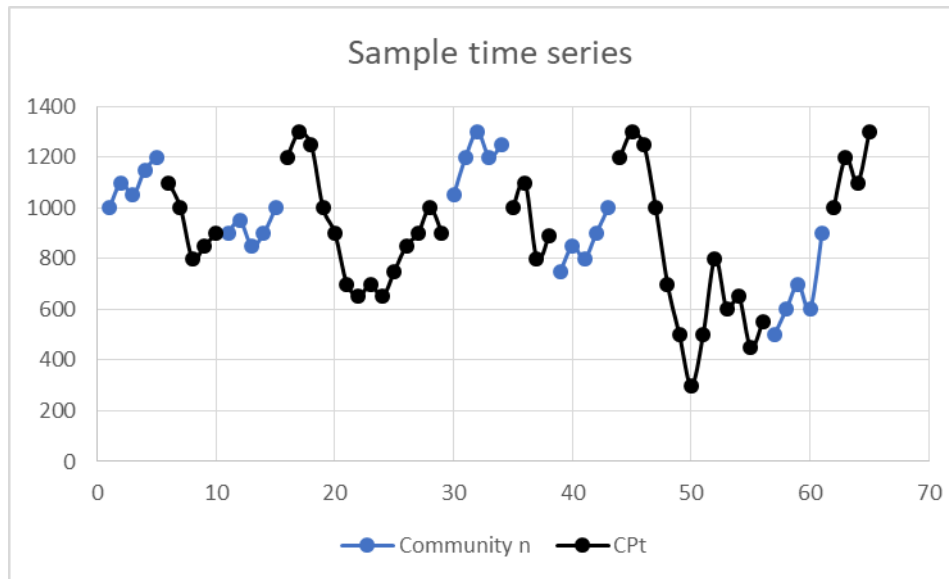


[2] Clauset, A., Newman, M.E. and Moore, C., 2004. Finding community structure in very large networks. Physical review E, 70(6), p.066111.

MÉTODO PROPOSTO

❑ Classificação das comunidades

❑ Retorno médio das ocorrências da comunidade da série original



$$\Delta cp_1 = 0.20$$

$$\Delta cp_2 = 0.11$$

$$\Delta cp_3 = 0.19$$

$$\Delta cp_4 = 0.33$$

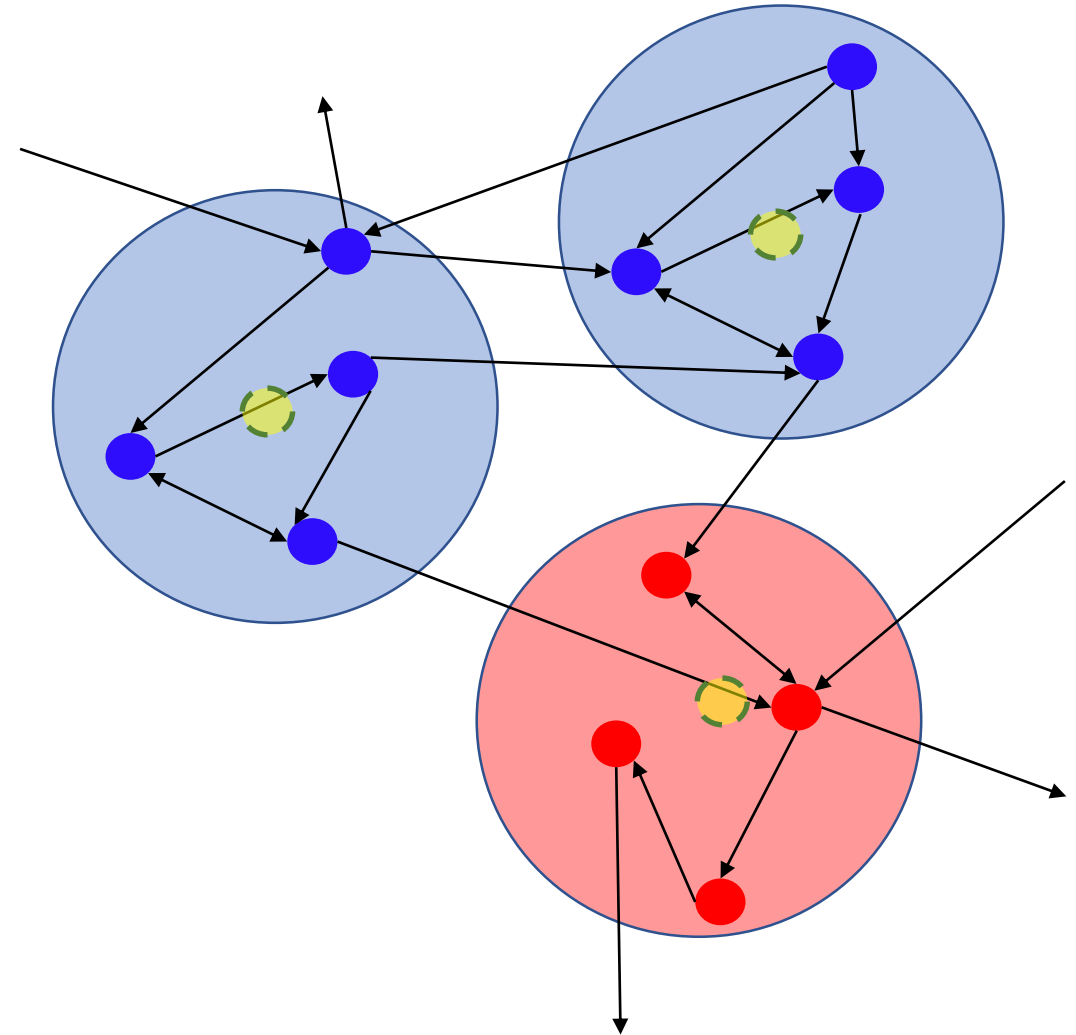
$$\Delta cp_5 = 0.80$$

$$\Delta cp_{avg} = 0.32$$

MÉTODO PROPOSTO

- ❑ Previsão de tendência
 - ❑ Tendência é gerada pelo vértice característico de cada comunidade
 - ❑ Vértice característico = centro de gravidade da comunidade, ponderado pelo grau
 - ❑ Similar ao kNN(1), mas considera apenas vértice mais importante de cada comunidade

$$C_c = \frac{\sum_1^n k_n * V_n}{n}$$



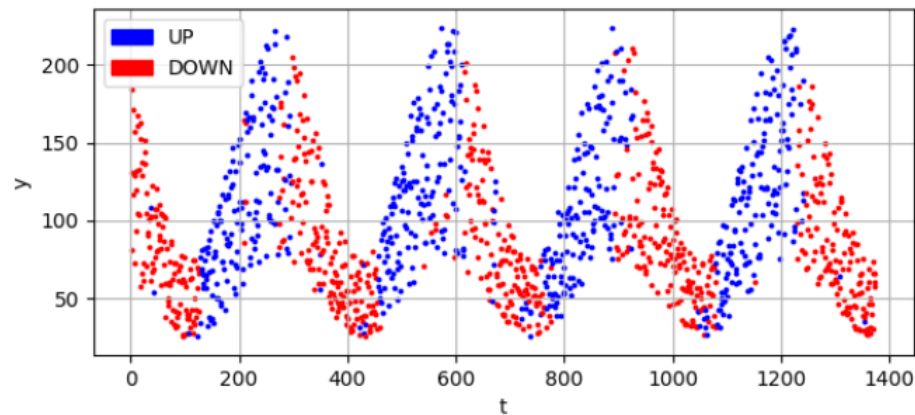
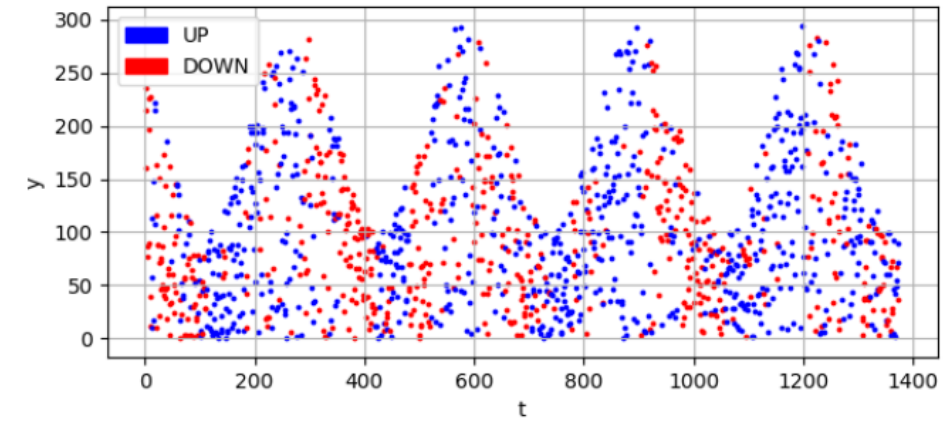
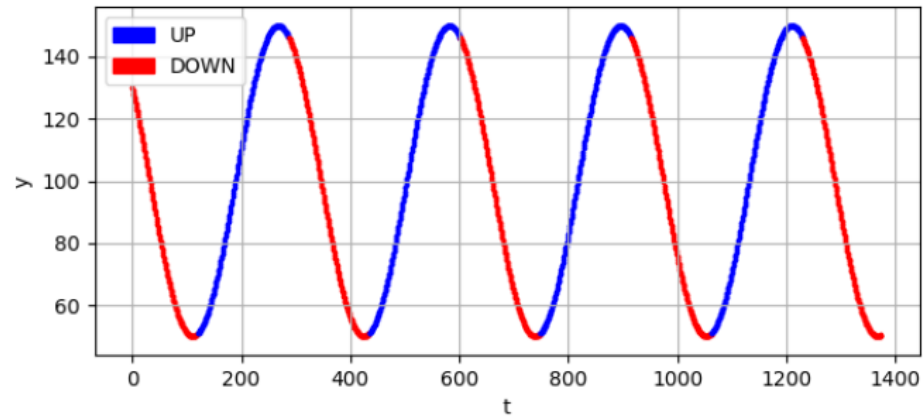
MÉTODO PROPOSTO

☐ Previsão de tendência - método



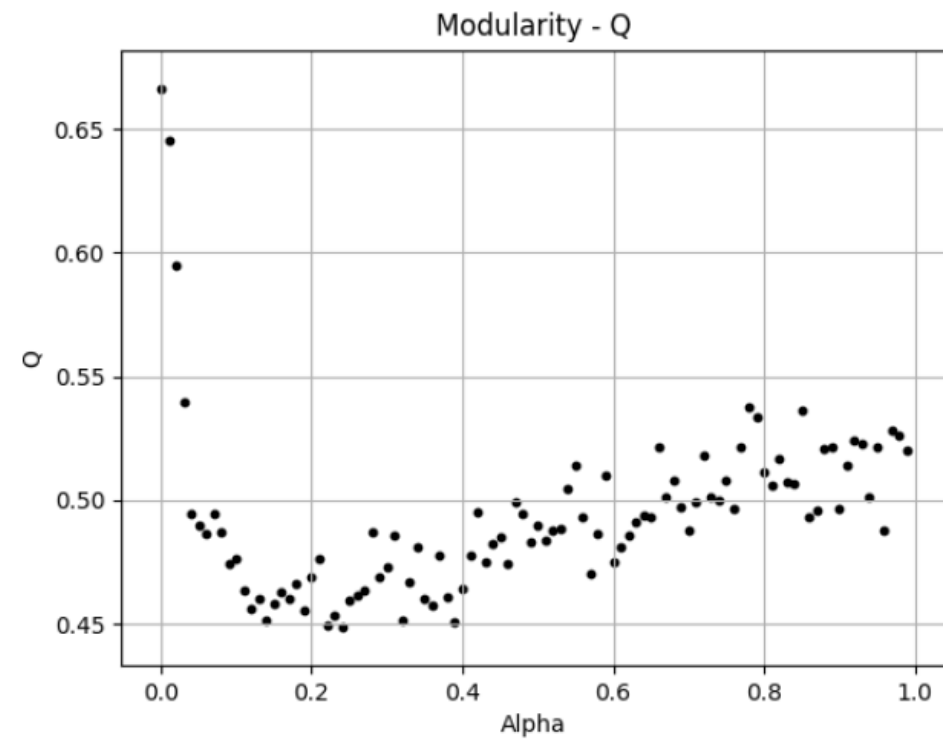
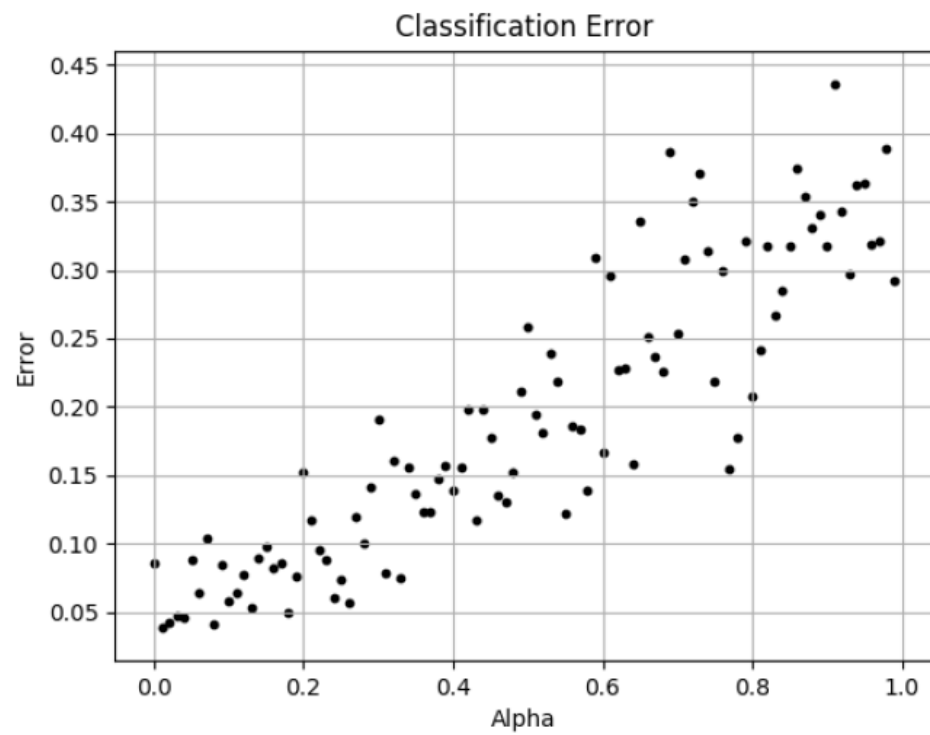
RESULTADOS

☐ Curva artificial - classificação



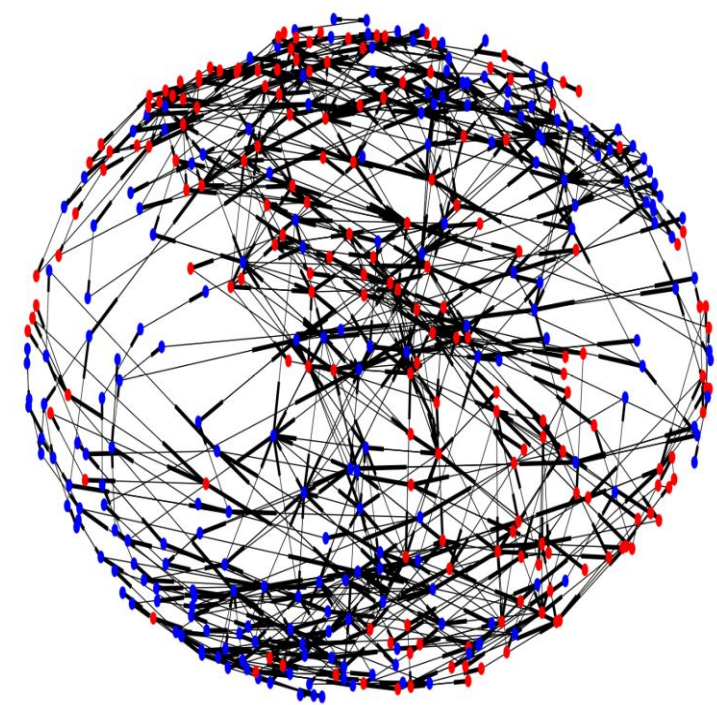
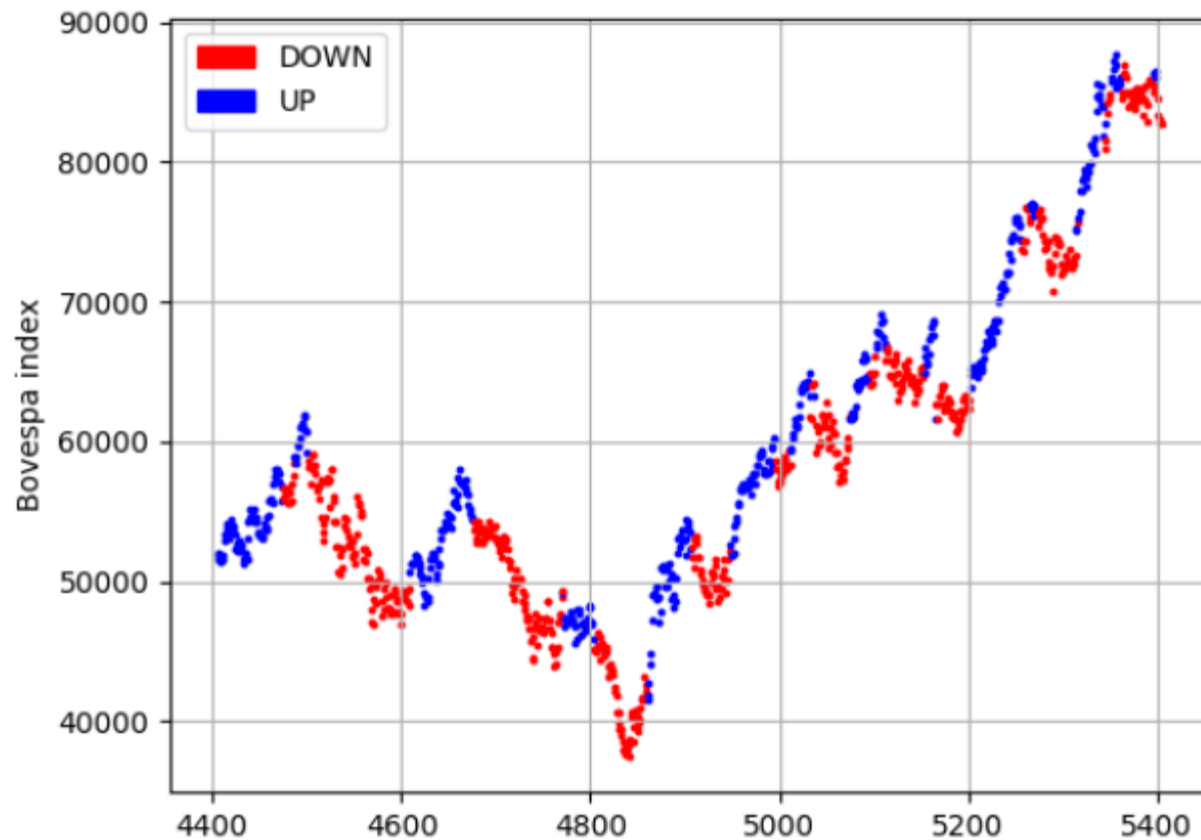
RESULTADOS

☐ Curva artificial - classificação



RESULTADOS

☐ Índice Bovespa (20 anos) - classificação



RESULTADOS

- ❑ Índice Bovespa (20 anos) – classificação
- ❑ Como avaliar se é uma boa classificação?

```
##### RETORNO HISTORICO #####
```

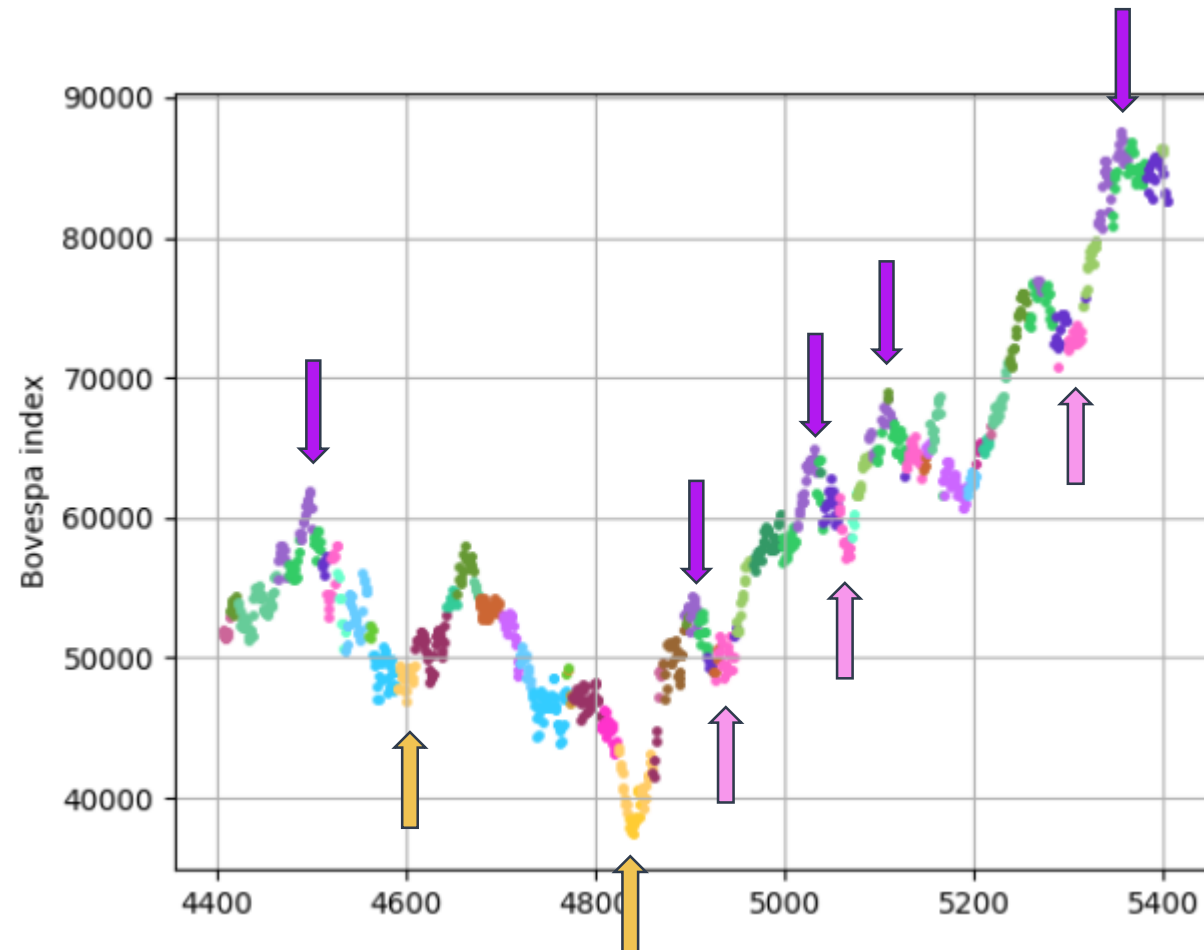
YEAR	\$\$TOTAL\$\$	TAXA	BLS	IBOV
1998	\$529,588.00	159.2	0.5	-25.7
1999	\$898,160.00	69.6	0.0	90.5
2000	\$2,425,451.00	170.0	0.4	6.7
2001	\$4,217,849.00	106.3	0.0	-6.8
2002	\$7,285,395.00	72.7	0.0	-16.6
2003	\$11,895,459.00	63.3	0.1	104.4
2004	\$26,680,821.00	124.3	0.4	26.8
2005	\$41,810,978.00	56.7	0.4	32.8
2006	\$63,323,989.00	51.5	0.4	35.1
2007	\$115,918,793.00	83.1	0.8	51.9
2008	\$263,576,523.00	128.8	0.1	-40.8
2009	\$363,676,333.00	38.0	0.4	86.7
2010	\$622,870,055.00	71.3	0.8	2.6
2011	\$865,578,575.00	39.0	0.7	-17.6
2012	\$1,289,710,880.00	55.2	1.2	4.3
2013	\$1,839,520,291.00	41.8	1.2	-11.3
2014	\$2,924,102,707.00	59.0	0.0	-2.7
2015	\$4,153,851,708.00	42.1	1.2	-15.2
2016	\$7,822,737,457.00	88.3	1.2	28.8
2017	\$11,176,222,785.00	48.2	0.0	21.6
2018	\$12,889,244,313.00	15.3	0.0	13.1

TAXA TTL: 66.80 a.a
BLS MEDIO: 0.10
INDICE QUALIDADE: 655.09

Não é backtest

RESULTADOS

☐ Índice Bovespa (20 anos) – classificação por comunidade



RESULTADOS

❑ Índice Bovespa – classificação por comunidade (tabela parcial)

Total Observations (Series length): 5405 Network Modularity (Q): 0.543 Total node count: 910 Total community count: 34							
Community ID	Node Count	Day Count	% Part	Occurrences	%Δcp	Trend	%Com_Change
A	21	193	3.6	5	16.94	UP	60.0
B	22	460	8.5	18	7.93	UP	38.9
C	59	119	2.2	3	7.43	UP	0.0
D	36	315	5.8	25	4.91	UP	36.0
E	22	53	1.0	4	4.0	UP	0.0
F	53	263	4.9	15	3.79	UP	6.7
G	38	151	2.8	12	3.24	UP	8.3
H	11	18	0.3	2	3.17	UP	0.0
I	44	300	5.6	18	2.31	UP	50
J	31	86	1.6	2	2.23	UP	0.0

❑ Índice Bovespa – previsão

Method	Precision	Recall	F-measure	Accuracy
Proposed	0.83 ± 0.27	0.92 ± 0.11	0.89 ± 0.13	0.93 ± 0.07
DET	0.78 ± 0.14	0.80 ± 0.11	0.78 ± 0.08	0.79 ± 0.07
NB	0.76 ± 0.16	0.77 ± 0.13	0.75 ± 0.10	0.76 ± 0.09
MLP	0.72 ± 0.16	0.74 ± 0.15	0.71 ± 0.11	0.73 ± 0.10

PRÓXIMOS PASSOS

- Desenvolvimento do modelo de classificação:
 - Utilização de novos métodos de clusterização (como competição de partículas – classificação fuzzy)
- Desenvolvimento do modelo de previsão:
 - Novos modelos baseados em métricas da rede
 - Utilização de métricas da rede com classificadores existentes (DT, MLP, etc)
- Inclusão de novos datasets:
 - Novas curvas artificiais
 - Novas series reais (S&P500, Forex)
- Proposta de novo framework:
 - Análise de multiplas series através de redes multicamadas

Obrigado!

Leandro Anghinoni

anghinoni@usp.br