

Dinâmica da Reciprocidade Periférica em uma Rede de Citações Acadêmicas

Ana Carolina Wagner Gouveia de Barros

Orientador: Walter Wagner Carvalho Sande

Dezembro | 2016

INTRODUÇÃO – Redes de Citações Acadêmicas

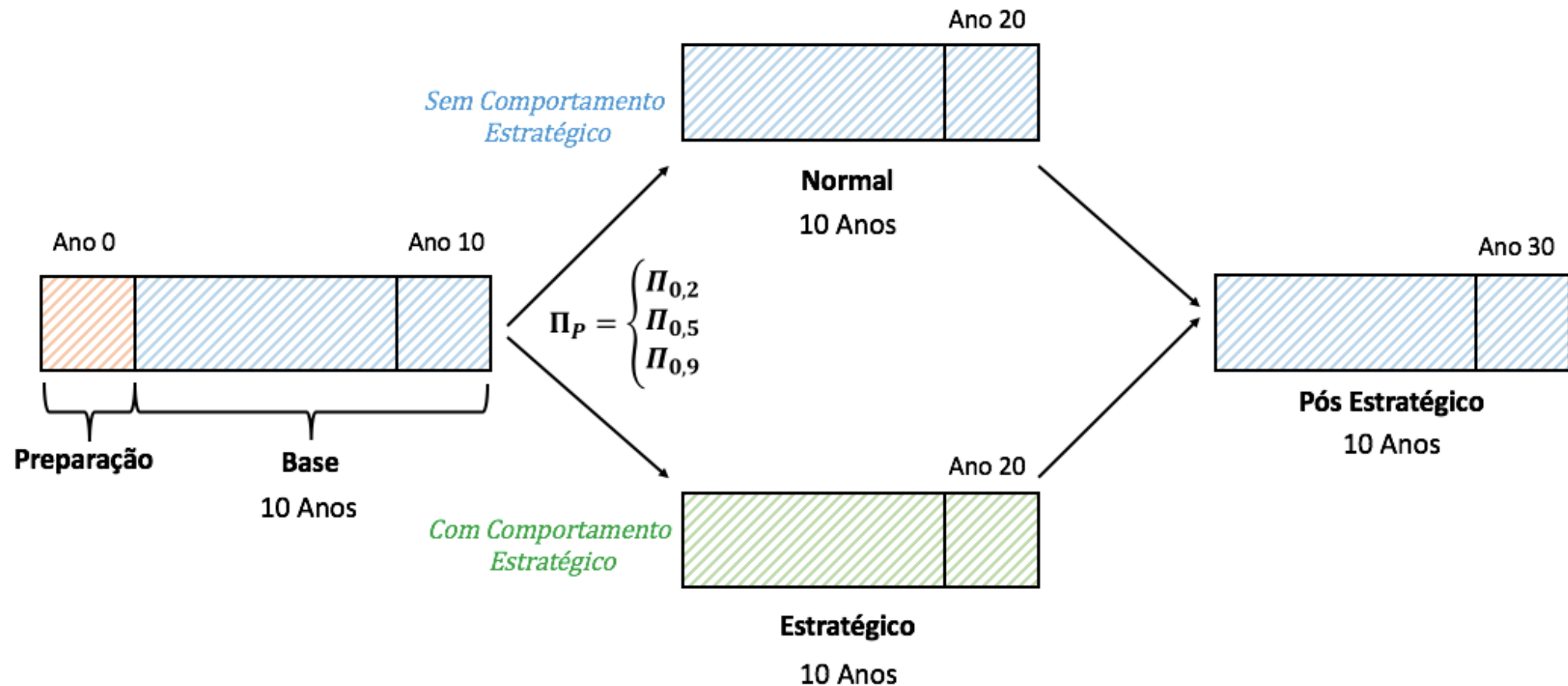
- **Redes direcionadas.**
- **Em uma rede é fundamental medir e avaliar a importância dos nós que a compõe.**
 - Fator de Impacto: mede a centralidade de grau dos periódicos. Calculado pela *Thomson Reuters*, trata-se de um análise quantitativa das citações. Reputação acadêmica baseada na contagem de citações.
- **O surgimento de novos nós na rede segue uma Lei de Potências e é aderente ao princípio da Ligação Preferencial.**
- **Presença de homofilia, o que proporciona o surgimento de uma estrutura de centro-periferia.**
- **Como novos artigos podem ser citados? Um novo artigo tem um certo *fitness* que influencia sua capacidade de atrair ligações.**
- **Em uma rede social há diferença entre popularidade e prestígio.**
 - Centralidade por autovetor – citações provenientes de periódicos de maior reputação.

INTRODUÇÃO – Objetivos

- **Desafios dos nós periféricos em alcançar maiores níveis de centralidade em redes sociais fortemente influenciadas pela ligação preferencial.**
 - Matthew Effect: à medida que a rede cresce, novos artigos tendem a citar artigos de grande prestígio acadêmico (maior grau de entrada).
- **Quais estratégias os nós periféricos podem adotar a fim de aumentar seu poder atrativo?**
 - Comportamento Estratégico: aumento da reciprocidade de citações entre os artigos publicados pelos periódicos periféricos.
 - **Hipótese 1**: incremento da centralidade em termos de citações recebidas (grau de entrada) (FI2/FI5).
 - **Hipótese 2**: não reflete em um aumento de centralidade medida por autovetor (EIG)
 - **Hipótese 3**: o índice de reciprocidade de um periódico periférico consegue prever sua centralidade de grau (R).
- **Objetivos:**
 - Mudança no processo de criação das redes. Uma rede mais fiel a realidade das citações acadêmicas.
 - Analisar como os periódicos periféricos incorporam os ganhos de centralidade obtidos a partir da adoção do Comportamento Estratégico.

METODOLOGIA – Aspectos Gerais da Simulação Original

- Foram geradas duas redes similares às redes acadêmicas: rede de artigos e rede de periódicos.
- Ambiente de programação: *Matlab*
- Cada rodada de simulação foi repetida três vezes (garantia de aleatoriedade).



- **Período de Preparação: criação das duas redes**
 - Criação de uma base com $NP = 400$ periódicos aos quais foram atribuídos alguns parâmetros: quantidade inicial de artigos, quantidade de artigos publicada por ano e média de *fitness* dos artigos publicados.
 - Para cada um desses artigos também foram atribuídos parâmetros: ano, periódico, fitness e número de citações.
 - Nesse período, a probabilidade de um artigo a ser citado por i , é proporcional somente ao seu *fitness*.

$$P(a:alvo) = \frac{\varphi_a}{\sum_i \varphi_i}$$

Parâmetro	Descrição	Valor
NP	Quantidade total de periódicos	400
\overline{NA}_{inic}	Quantidade média de artigos dos periódicos no início da simulação	200
\overline{NA}	Quantidade média de artigos publicados por cada periódico a cada ano	25
\overline{NCIT}	Quantidade média de citações de cada artigo	20
Φ_{max}	Limite da faixa de fitness dos artigos publicados	100

- **Período Base: início da dinâmica de crescimento**

- A cada ano, os *NP* periódicos publicam novos artigos, aos quais são atribuídos os parâmetros já citados anteriormente, que, por sua vez, citam os artigos já existentes na rede.
- Agora, 50% dos casos passam a seguir o modelo de Bianconi-Barabási (2001): *fitness* ponderado.

$$P(a:alvo) = \frac{\varphi_a k_a^{in}}{\sum_i \varphi_i k_i^{in}} ; \quad k_i^{in} = \sum_j cit_{j \rightarrow i}$$

- Ao final do período,
 - Periódicos Periféricos: periódicos que ao final do Período Base, apresentam baixa centralidade, medida pelo valor de **FI2**.
 - Periódicos Estratégicos: aqueles que irão apresentar Comportamento Estratégico. De acordo com probabilidades $\Pi_P = [0,2; 0,5; 0,9]$ são criados três grupos distintos.
 - Duas possibilidades: seguir para o **Período Normal**, que possui a mesma dinâmica ou seguir para o **Período Estratégico**.

METODOLOGIA – Dinâmica da Simulação Original

- **Período Estratégico: início do Comportamento Estratégico**
 - O objetivo é fugir da dinâmica da ligação preferencial.
 - Cada Periódico Estratégico ($\Pi_P = [0,2; 0,5; 0,9]$) possui um grupo de Artigos Estratégicos ($\Pi_A = [0,1; 0,5; 0,9]$) que por sua vez possui uma parcela de Citações Estratégicas ($\Pi_C = [0,1; 0,5; 0,9]$). A ideia é direcionar essas citações apenas para artigos classificados como sendo estratégicos.
 - Periódicos Normais seguem a dinâmica de crescimento de rede idêntica ao Período Base.
- **Período Pós-Estratégico: fim do Comportamento Estratégico**
- **Período Normal: comparação dos resultados**
 - Os três grupos de Periódicos Estratégicos criados ao final do Período Base não desenvolvem Comportamento Estratégico ($\Pi_A, \Pi_C = 0$).
- **A cada ano, durante todo o processo de simulação, após a publicação dos artigos e a alocação de suas citações, atualizam-se as ligações entre os periódicos em que tais artigos foram publicados, conduzindo assim, a dinâmica de crescimento da rede de citações entre periódicos.**

METODOLOGIA – Índices Analisados

- Ao final de cada ano, são calculadas e atualizadas as informações referentes ao grau de entrada, *fitness* médio dos artigos publicados e fator de impacto de cada periódico.
- Ao final de cada período de simulação, Período Base (ano=10), Período Normal ou Estratégico (ano=20) e Período Pós-Estratégico (ano=30) são calculados os índices de reciprocidade de cada periódico e a centralidade medida por autovetor (Eigenfactor).

- Fator de Impacto (mensuração da centralidade de grau dos periódicos)

$$FI2 = \frac{k2^{in}}{na} \qquad FI5 = \frac{k5^{in}}{na}$$

- Reciprocidade

$$\text{Citações Recíprocas: } r_{P_1 \rightarrow P_2} = \min(cit_{P_1 \rightarrow P_2}; cit_{P_2 \rightarrow P_1})$$

$$R_i = \frac{r_i}{k_i^{out}}$$

- Eigenfactor (mensuração do prestígio dos periódicos)
 - Um nó importante é aquele que está conectado a outros nós também importantes.

METODOLOGIA – Simulação Proposta

- **Alteração do Período de Preparação e de como ocorre o processo de submissão e publicação dos artigos durante a dinâmica de crescimento da rede.**
- **Simulação Original:**
 - Subordinação da geração dos artigos à existência dos periódicos.
 - O *fitness* dos artigos gerados foi determinado probabilisticamente pelo *fitness* do periódico, que é constante.
 - 1 rodada de Preparação.
- **Simulação Proposta:**
 - Incorporar o fato de que novos artigos estão sempre sendo lançados e que cada pesquisador tem seus periódicos de interesse, ocorrendo um *matching* entre artigos e periódicos.
 - Atualização do *fitness* dos periódicos conforme o crescimento da rede.
 - 10 rodadas de Preparação – estabilização da rede.

METODOLOGIA – Simulação Proposta

- **A cada ano,**
 - Os periódicos já possuem uma reputação inicial, herdada da média do fitness de seus artigos publicados em anos anteriores.
 - **Produção dos Artigos**
 - São produzidos NA_{prod} artigos com *fitness* a partir de uma distribuição Normal (independentemente dos periódicos).
 - **Submissão dos Artigos**
 - Os artigos são preferencialmente enviados aos periódicos que têm reputação semelhante ao seu *fitness*.
 - **Publicação dos Artigos**
 - Os periódicos fazem uma seleção dos artigos que serão publicados. É determinado que cada periódico escolhe os NA_{pub} melhores artigos (de maior *fitness*) que foram submetidos.
 - Se poucos artigos são publicados pelo periódico, ele morre (torna-se inativo). Quando muitos artigos são rejeitados por um periódico, há a criação de um novo periódico de igual reputação.
 - Com a publicação dos novos artigos, os periódicos atualizam sua reputação com base na média do fitness dos artigos publicados.

METODOLOGIA – Simulação Proposta

- O Período de Preparação inicia-se com apenas um periódico ativo.

Parâmetro	Descrição	Valor
NP_{max}	Quantidade máxima de periódicos	400
σ_{Φ}	Desvio padrão do <i>fitness</i> dos artigos produzidos	100
NA_{prod}	Quantidade de artigos produzidos por ano	12.000
NA_{pub}	Quantidade máxima de artigos publicados por cada periódico a cada ano	30
\overline{NCIT}	Quantidade média de citações de cada artigo	20

- A rede termina com 400 periódicos, mas nem todos encontram-se ativos.

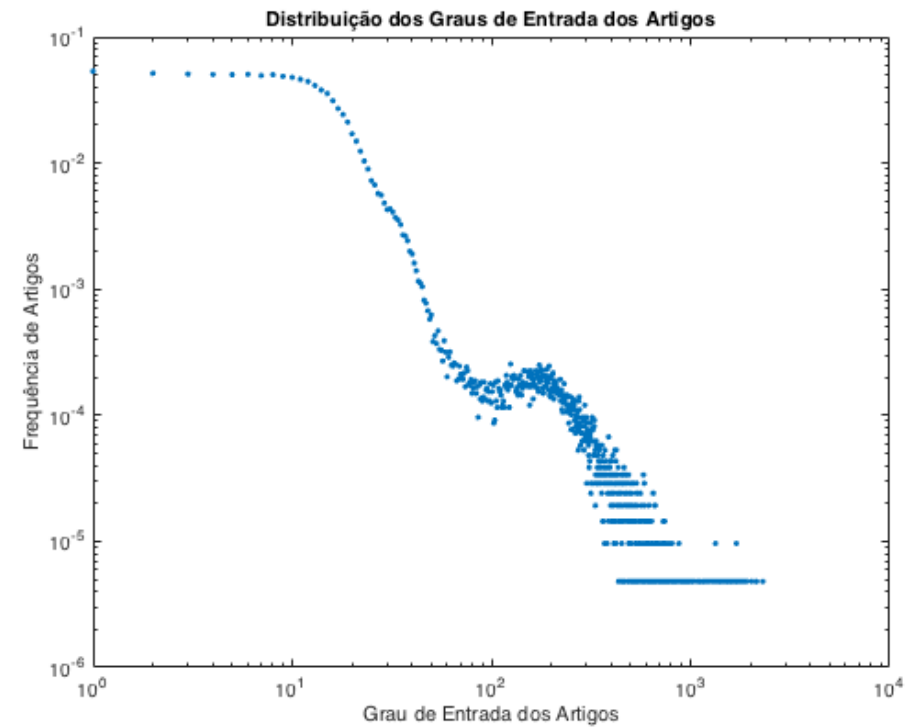
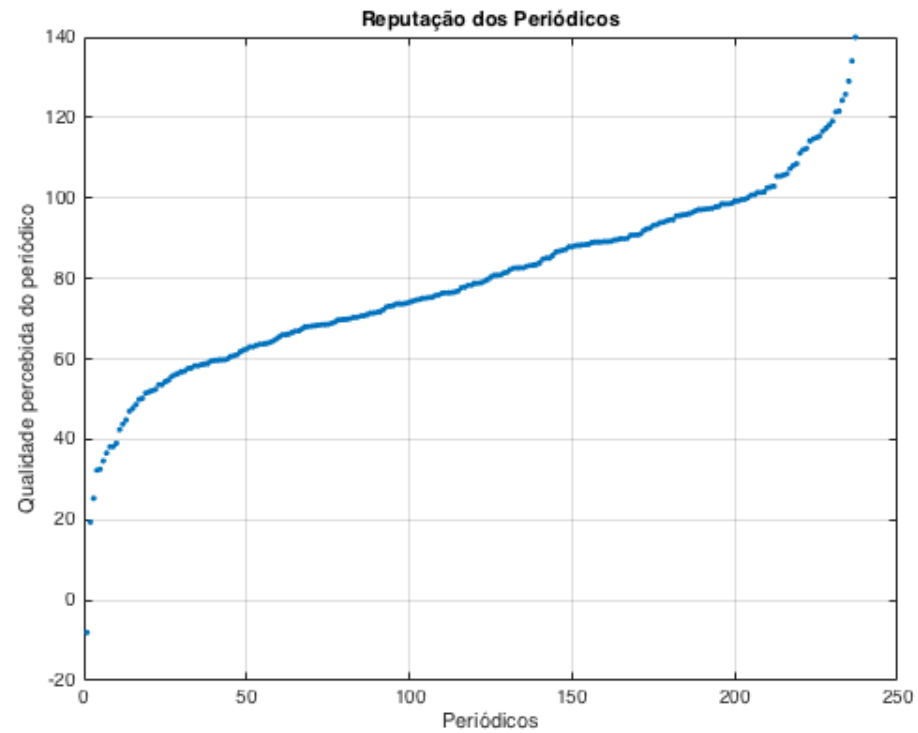
RESULTADOS – Validação dos Resultados

- O tempo de simulação total em cada rodada cresce exponencialmente com o tamanho da rede.
- O processo inteiro é muito custoso computacionalmente.
- Independência do tamanho da rede: Teste Kolmogorov-Smirnov (K-S)

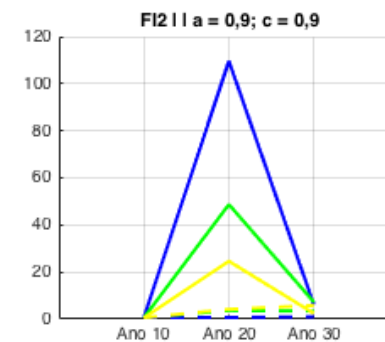
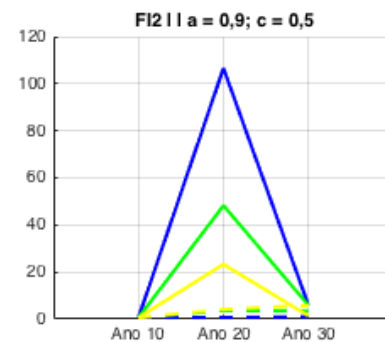
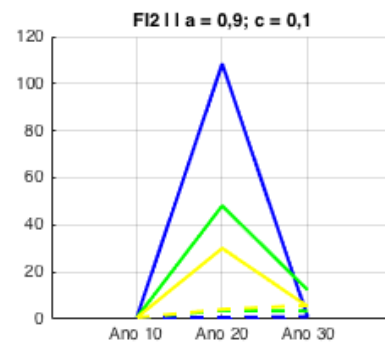
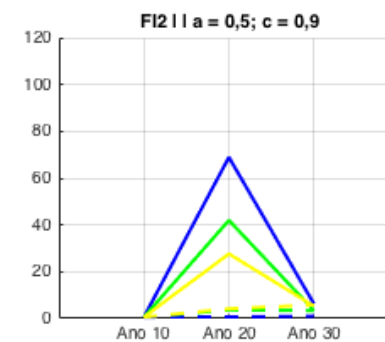
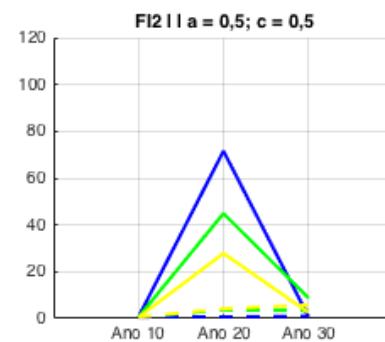
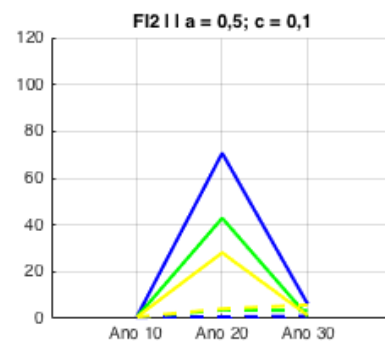
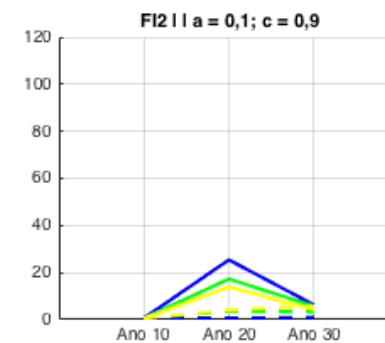
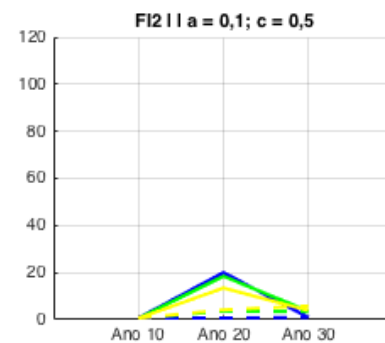
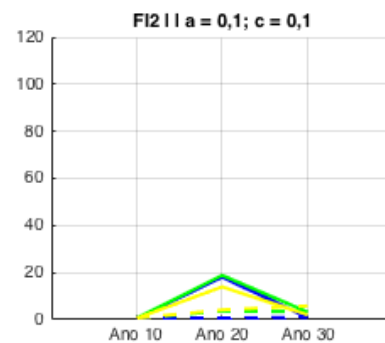
$$\begin{cases} H_0 : \text{As distribuições de FI5 são iguais} \\ H_a : \text{As distribuições de FI5 são diferentes} \end{cases}$$

- Com um nível de significância de 5%, em apenas 4,27% (210) dos 4.450 testes realizados foi recusada a hipótese nula. Concluindo-se assim que a premissa testada é válida, ou seja, as redes geradas ao final do Período Base são estacionárias.
- Ao todo, foram geradas 91 rodadas:
 - 1 rodada do Período Base, 3 repetições das 3 rodadas do Período Normal (9 rodadas) e 3 repetições das 27 rodadas do Período Estratégico (81 rodadas).

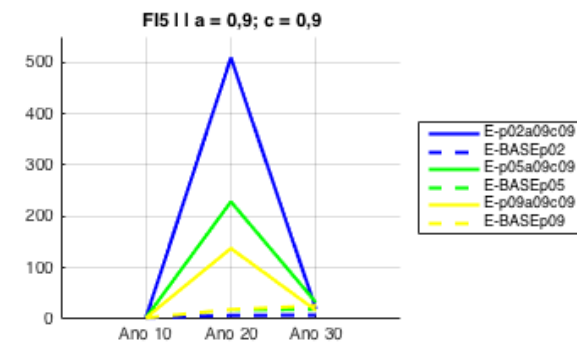
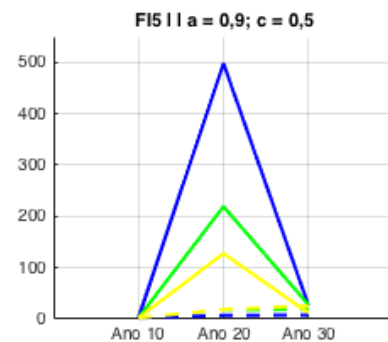
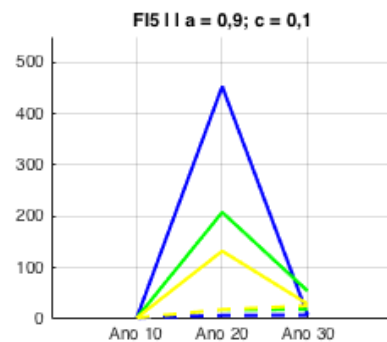
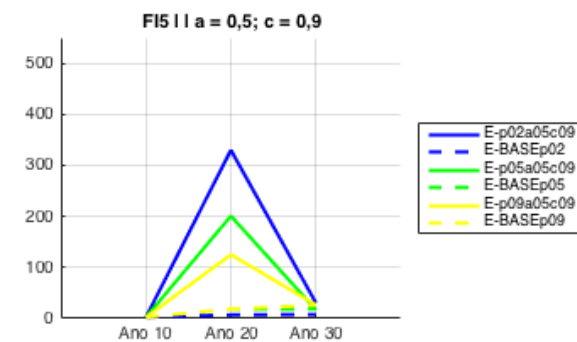
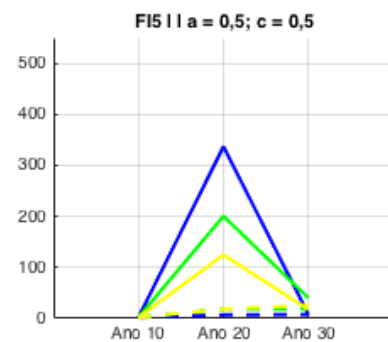
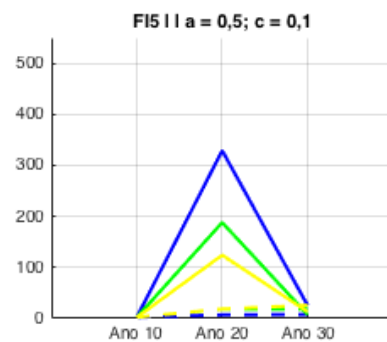
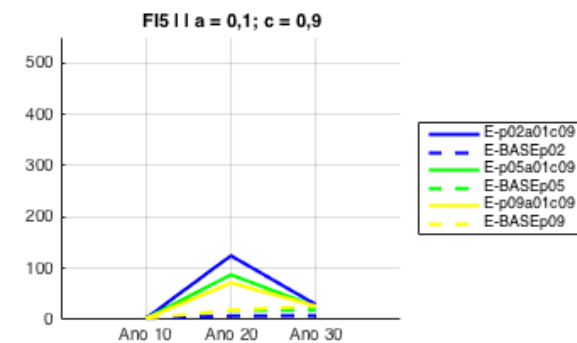
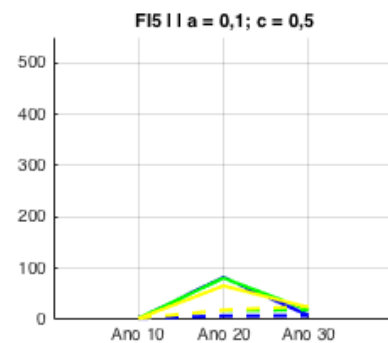
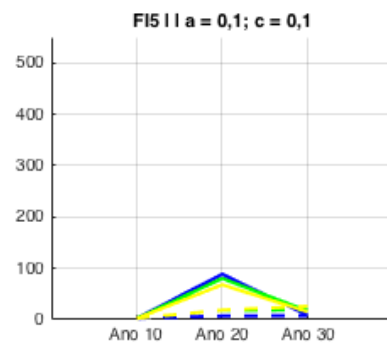
RESULTADOS



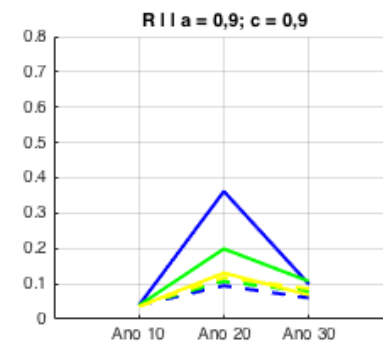
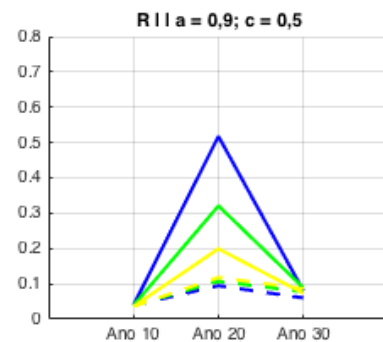
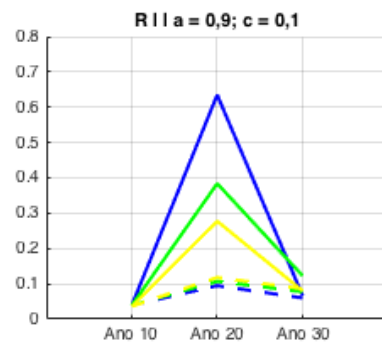
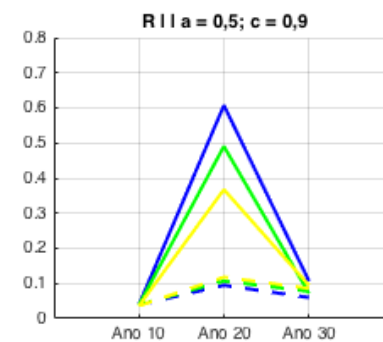
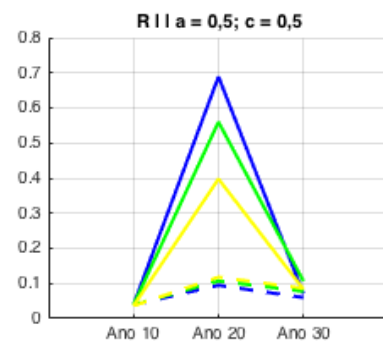
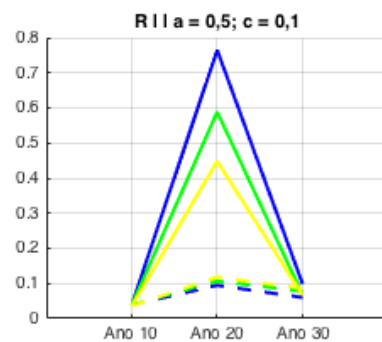
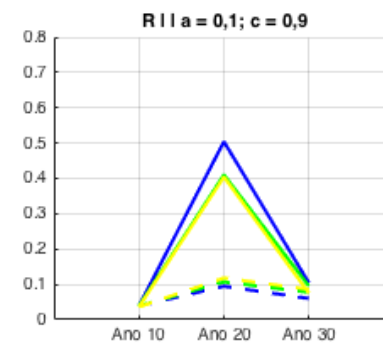
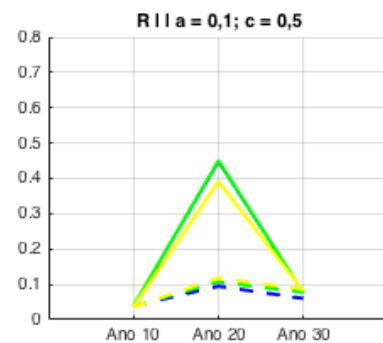
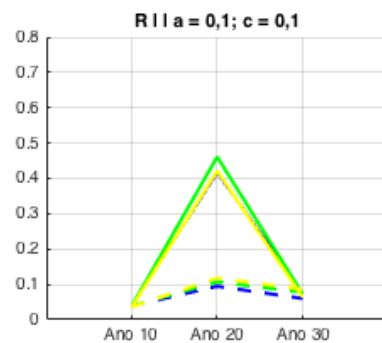
RESULTADOS – FI2



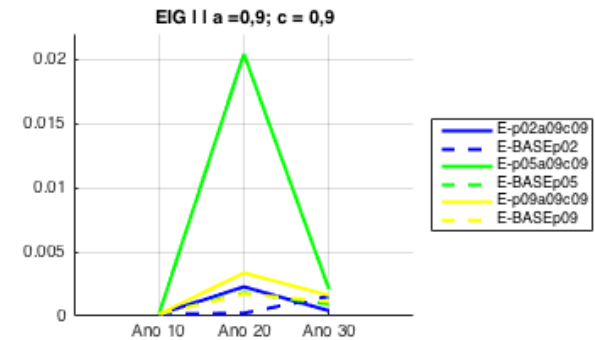
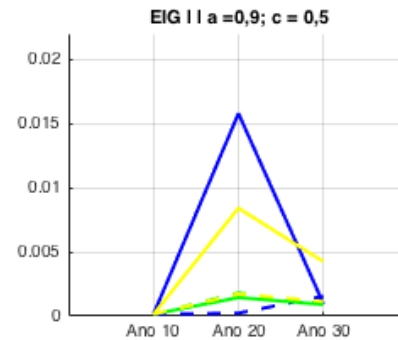
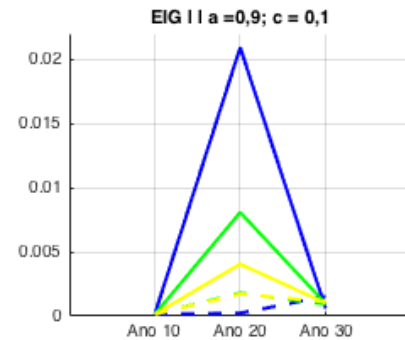
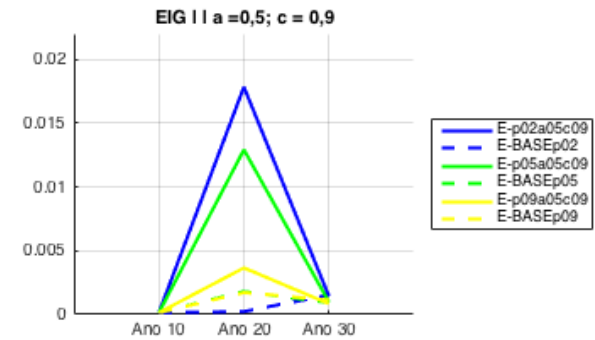
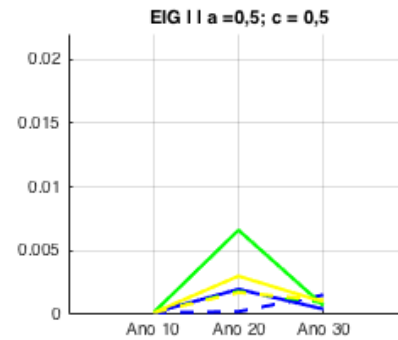
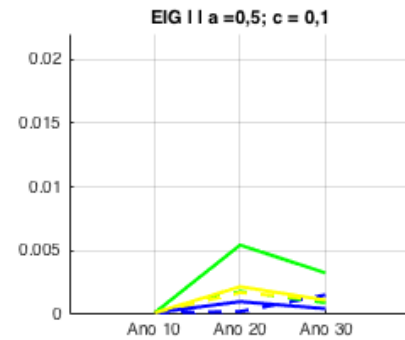
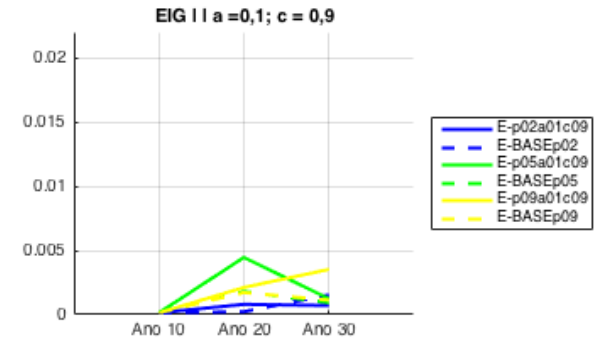
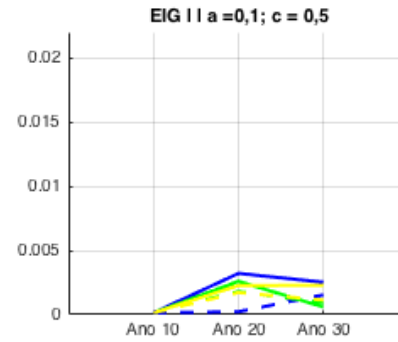
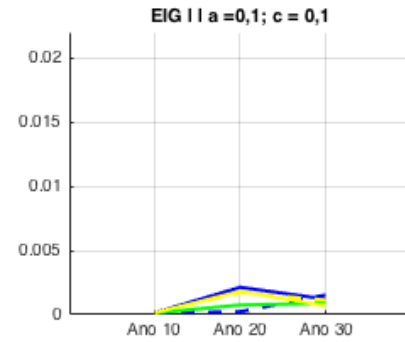
RESULTADOS – FI5



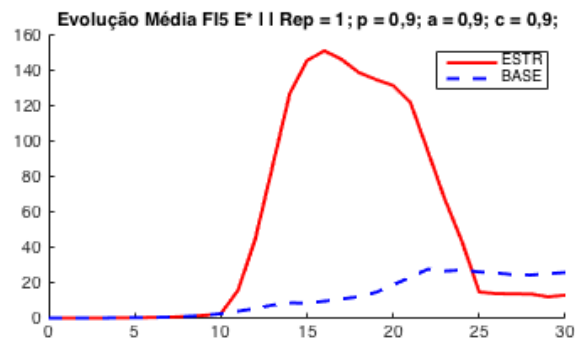
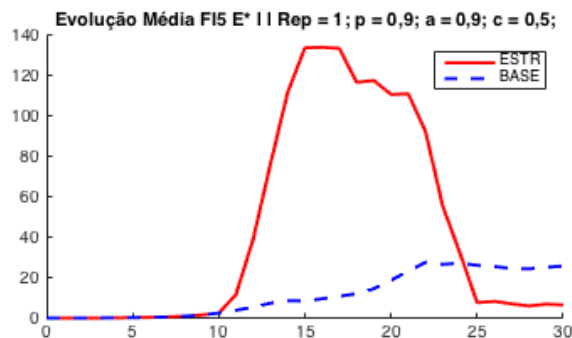
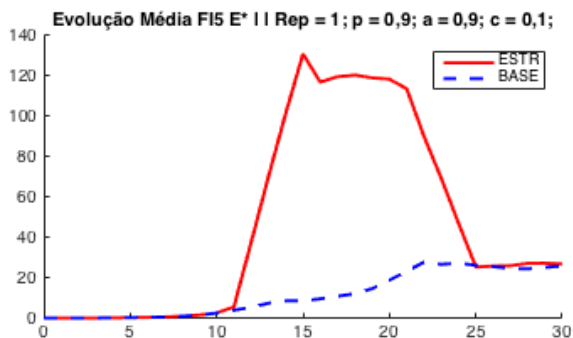
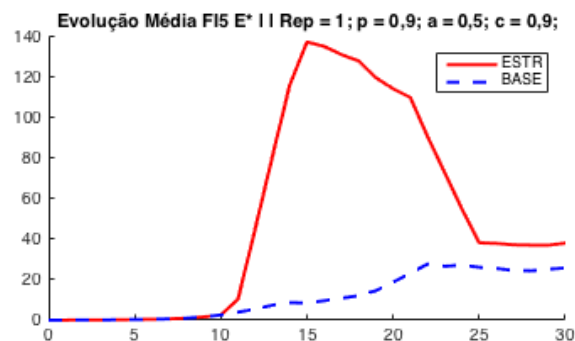
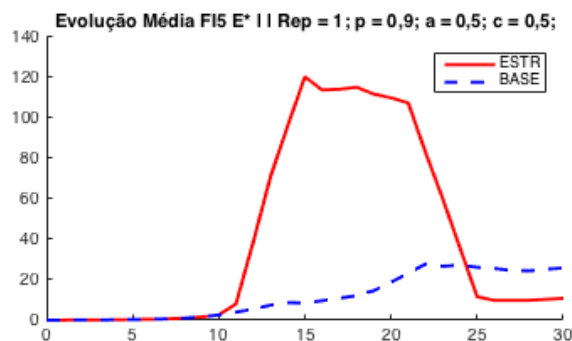
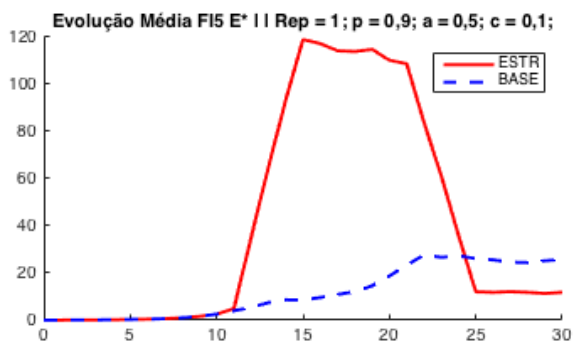
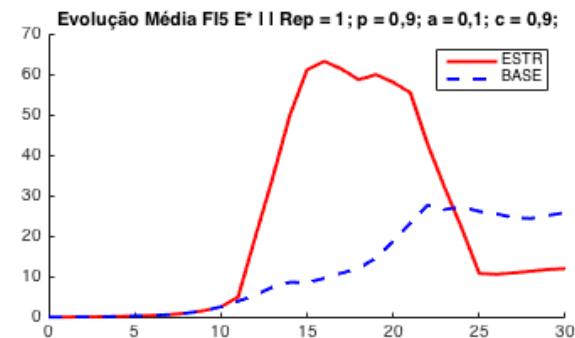
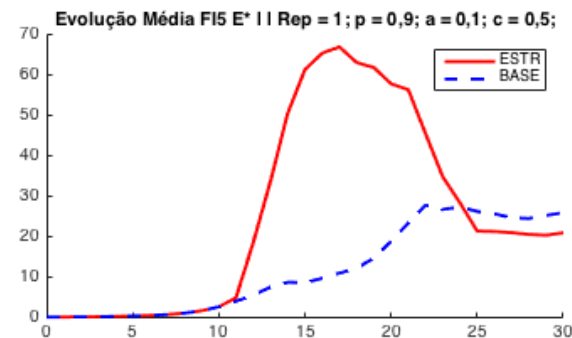
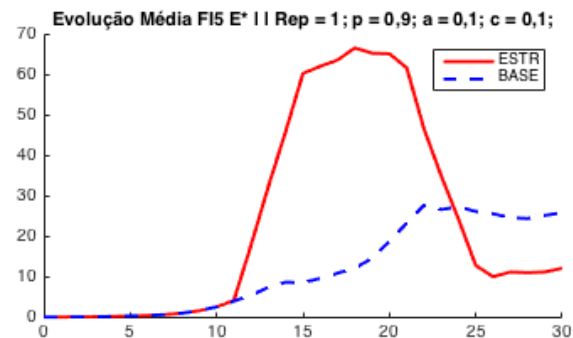
RESULTADOS - R



RESULTADOS - EIG



RESULTADOS – Evolução Média FI5 E*



RESULTADOS – Tabela de Medidas – FI2

			$\Pi p = 0,2$			$\Pi p = 0,5$			$\Pi p = 0,9$		
			Ano 10	Ano 20	Ano 30	Ano 10	Ano 20	Ano 30	Ano 10	Ano 20	Ano 30
Rodadas (repetições)			9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)
N Válidos			10800	10800	10800	10800	10800	10800	10800	10800	10800
N Periódicos / rodada			400	400	400	400	400	400	400	400	400
N Periféricos / rodada			65	65	65	65	65	65	65	65	65
N Estratégicos / rodada			11	11	11	35	35	35	61	61	61
N Total Estratégicos Inativos			0	25	38	0	124	162	0	183	254
FI2	E*	Média	0,676	66,914	3,877	0,645	36,684	5,739	0,634	22,606	3,191
		Mediana	0,667	41,917	0,833	0,600	16,700	0,900	0,600	7,933	0,867
	N*	Média	0,620	3,467	5,448	0,611	1,405	4,497	0,567	0,792	2,527
		Mediana	0,567	0,733	0,800	0,550	0,700	0,867	0,483	0,633	0,763
	C*	Média	16,690	10,783	14,616	16,690	8,585	14,784	16,690	7,825	14,753
		Mediana	3,733	3,367	3,533	3,733	3,167	3,633	3,733	3,133	3,600
	Mínimo		0,000	0,000	0,000	0,000	0,000	0,033	0,000	0,000	0,000
	Máximo		316,533	251,900	269,800	316,533	226,933	262,033	316,533	225,800	327,000
	Q1		1,400	1,667	1,533	1,400	1,833	1,633	1,400	1,967	1,633
	Q2		2,830	2,967	2,900	2,830	3,100	3,000	2,830	3,200	3,000
	Q3		4,400	4,233	4,200	4,400	4,400	4,367	4,400	4,533	4,280

RESULTADOS – Tabela de Medidas – FI5

			$\Pi_p = 0,2$			$\Pi_p = 0,5$			$\Pi_p = 0,9$		
			Ano 10	Ano 20	Ano 30	Ano 10	Ano 20	Ano 30	Ano 10	Ano 20	Ano 30
Rodadas (repetições)			9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)
N Válidos			10800	10800	10800	10800	10800	10800	10800	10800	10800
N Periódicos / rodada			400	400	400	400	400	400	400	400	400
N Periféricos / rodada			65	65	65	65	65	65	65	65	65
N Estratégicos / rodada			11	11	11	35	35	35	61	61	61
N Total Estratégicos Inativos			0	25	38	0	124	162	0	183	254
FI5	E*	Média	2,482	306,946	18,501	2,593	166,158	26,874	2,500	109,076	20,221
		Mediana	1,300	196,683	8,200	1,167	81,000	7,900	1,167	44,967	8,033
	N*	Média	2,537	19,291	29,847	2,452	10,979	27,791	2,950	7,258	19,005
		Mediana	1,050	7,267	7,867	0,917	6,700	7,900	0,817	5,600	6,420
	C*	Média	79,982	60,049	76,731	79,982	51,124	76,599	79,982	47,370	78,310
		Mediana	30,867	32,167	33,700	30,867	30,983	33,267	30,867	29,700	33,667
	Mínimo		0,000	0,567	0,867	0,000	0,333	1,067	0,000	0,600	0,833
	Máximo		1211,233	1022,833	1031,667	1211,233	917,267	1023,133	1211,233	955,160	1239,208
	Q1		9,067	16,092	15,100	9,067	18,567	15,067	9,067	19,833	15,160
	Q2		23,233	28,667	28,067	23,233	29,933	27,333	23,233	30,067	27,833
	Q3		38,467	40,433	40,267	38,467	41,575	39,967	38,467	41,433	40,208

RESULTADOS – Tabela de Medidas – R

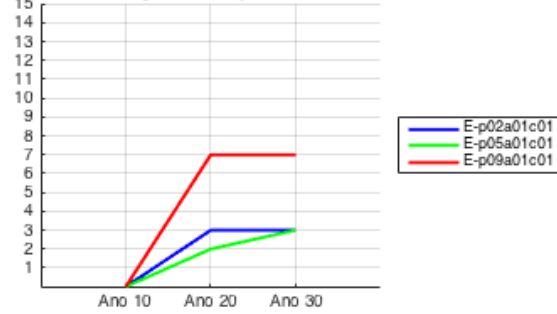
			$\Pi p = 0,2$			$\Pi p = 0,5$			$\Pi p = 0,9$		
			Ano 10	Ano 20	Ano 30	Ano 10	Ano 20	Ano 30	Ano 10	Ano 20	Ano 30
Rodadas (repetições)			9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)
N Válidos			10800	10800	10800	10800	10800	10800	10800	10800	10800
N Periódicos / rodada			400	400	400	400	400	400	400	400	400
N Periféricos / rodada			65	65	65	65	65	65	65	65	65
N Estratégicos / rodada			11	11	11	35	35	35	61	61	61
N TOTAL Estratégicos Inativos			0	25	38	0	124	162	0	183	254
R	E*	Média	0,035	0,549	0,071	0,034	0,427	0,071	0,032	0,332	0,058
		Mediana	0,035	0,549	0,071	0,034	0,427	0,071	0,032	0,332	0,058
	N*	Média	0,031	0,115	0,068	0,029	0,111	0,064	0,034	0,097	0,054
		Mediana	0,023	0,092	0,052	0,023	0,097	0,054	0,030	0,083	0,048
	C*	Média	0,274	0,346	0,290	0,274	0,350	0,290	0,272	0,346	0,287
		Mediana	0,258	0,329	0,260	0,258	0,333	0,261	0,253	0,328	0,254
	Mínimo		0,000	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Máximo		0,923	1,000	1,000	0,923	0,998	1,000	0,923	1,000	1,000
	Q1		0,089	0,188	0,087	0,089	0,219	0,087	0,090	0,227	0,085
	Q2		0,207	0,298	0,194	0,205	0,318	0,194	0,205	0,323	0,190
	Q3		0,291	0,381	0,327	0,291	0,408	0,329	0,290	0,418	0,321

RESULTADOS – Tabela de Medidas – EIG

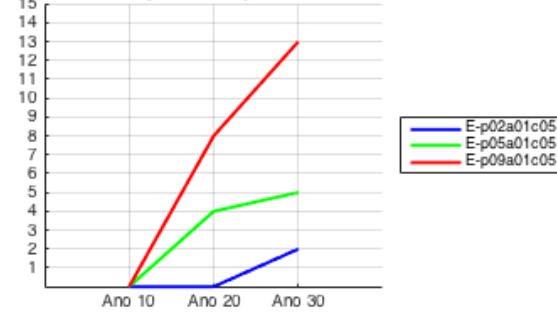
			$\Pi p = 0,2$			$\Pi p = 0,5$			$\Pi p = 0,9$		
			Ano 10	Ano 20	Ano 30	Ano 10	Ano 20	Ano 30	Ano 10	Ano 20	Ano 30
Rodadas (repetições)			9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)	9 (3)
N Válidos			10800	10800	10800	10800	10800	10800	10800	10800	10800
N Periódicos / rodada			400	400	400	400	400	400	400	400	400
N Periféricos / rodada			65	65	65	65	65	65	65	65	65
N Estratégicos / rodada			11	11	11	35	35	35	61	61	61
N TOTAL Estratégicos Inativos			0	25	38	0	124	162	0	183	254
EIG	E*	Média	1,47E-04	9,10E-04	3,89E-04	1,47E-04	4,82E-04	4,39E-04	1,47E-04	6,62E-04	4,50E-04
		Mediana	1,47E-04	9,10E-04	3,89E-04	1,47E-04	4,82E-04	4,39E-04	1,47E-04	6,62E-04	4,50E-04
	N*	Média	1,47E-04	2,27E-04	4,28E-04	1,47E-04	2,05E-04	4,81E-04	1,48E-04	2,35E-04	4,56E-04
		Mediana	1,46E-04	2,12E-04	3,12E-04	1,46E-04	2,06E-04	3,27E-04	1,46E-04	2,18E-04	3,33E-04
	C*	Média	2,26E-04	3,26E-04	7,09E-04	2,27E-04	2,93E-04	8,27E-04	2,26E-04	3,61E-04	8,78E-04
		Mediana	1,84E-04	2,79E-04	4,97E-04	1,83E-04	2,65E-04	5,27E-04	1,83E-04	2,90E-04	5,67E-04
	Mínimo		1,45E-04	1,20E-04	1,73E-04	1,45E-04	7,80E-05	2,00E-04	1,45E-04	1,51E-04	1,92E-04
	Máximo		1,35E-03	8,36E-03	1,22E-02	1,35E-03	6,88E-03	1,66E-02	1,35E-03	9,13E-02	1,48E-02
	Q1		1,53E-04	2,10E-04	3,18E-04	1,53E-04	2,00E-04	3,64E-04	1,53E-04	2,41E-04	3,73E-04
	Q2		1,73E-04	2,62E-04	4,32E-04	1,73E-04	2,58E-04	4,72E-04	1,73E-04	2,92E-04	4,98E-04
	Q3		1,95E-04	3,23E-04	5,93E-04	1,96E-04	3,25E-04	6,33E-04	1,95E-04	3,89E-04	6,75E-04

RESULTADOS – Periódicos Estratégicos Inativos

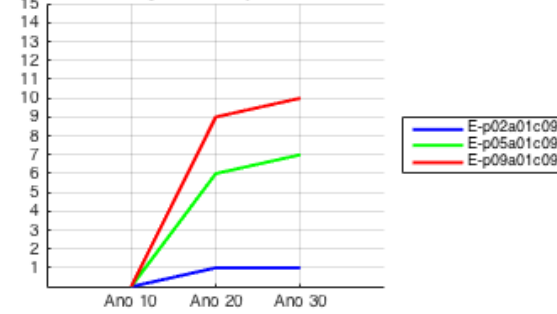
Inativos Estratégicos I | Rep = 1; a = 0,1; c = 0,1



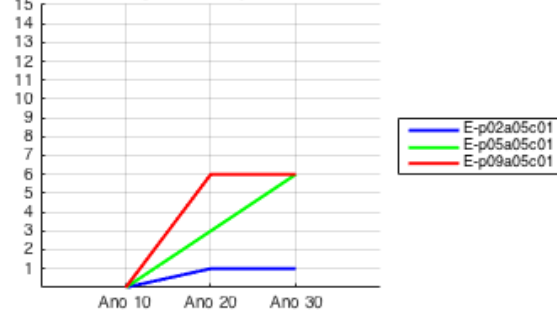
Inativos Estratégicos I | Rep = 1; a = 0,1; c = 0,5



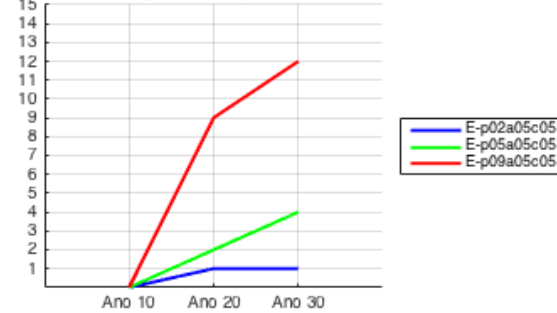
Inativos Estratégicos I | Rep = 1; a = 0,1; c = 0,9



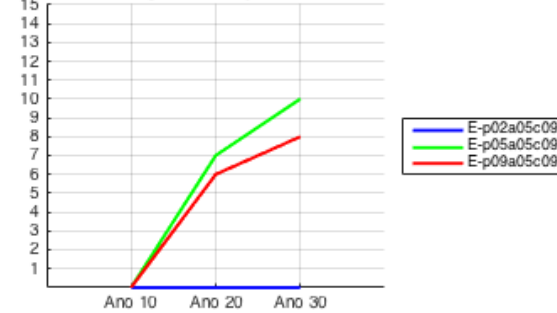
Inativos Estratégicos I | Rep = 1; a = 0,5; c = 0,1



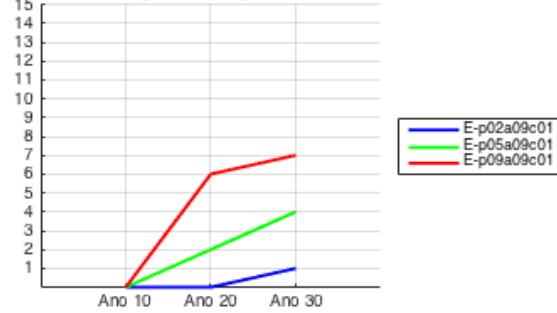
Inativos Estratégicos I | Rep = 1; a = 0,5; c = 0,5



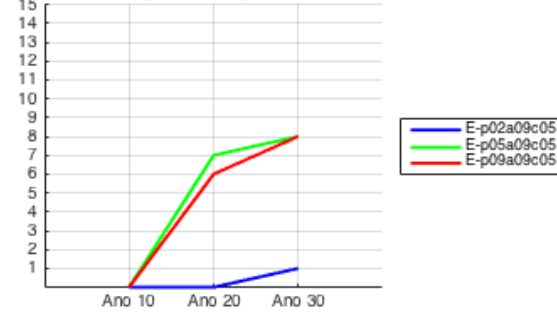
Inativos Estratégicos I | Rep = 1; a = 0,5; c = 0,9



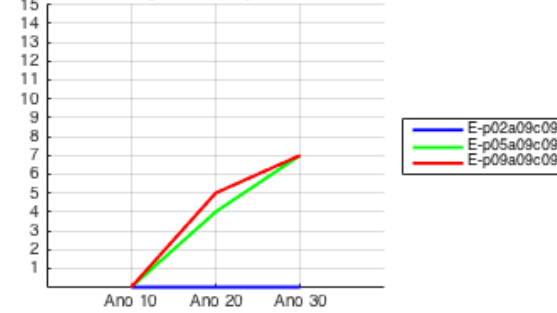
Inativos Estratégicos I | Rep = 1; a = 0,9; c = 0,1



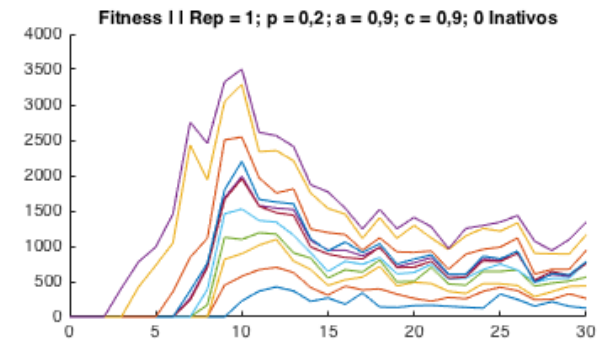
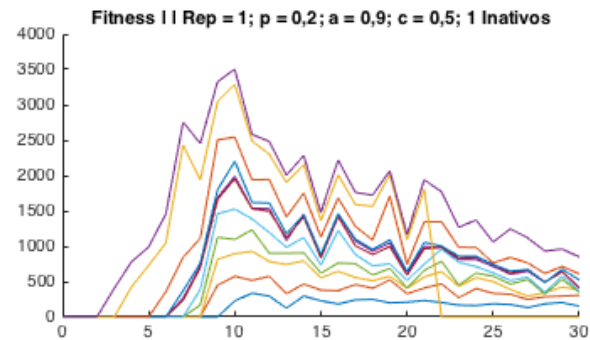
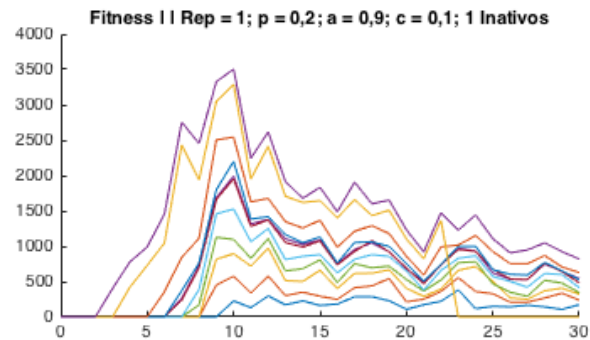
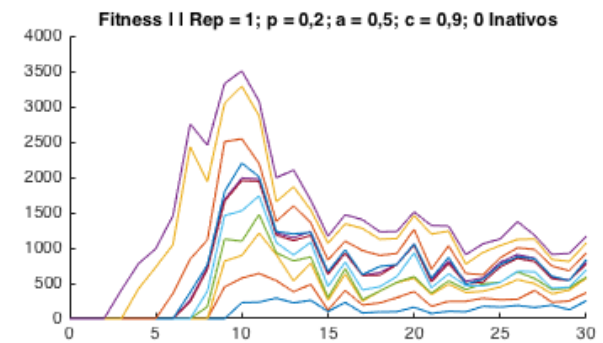
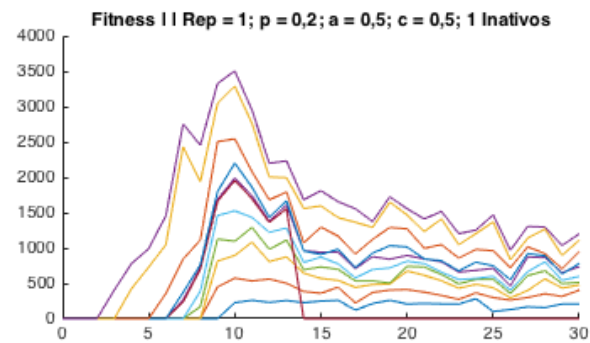
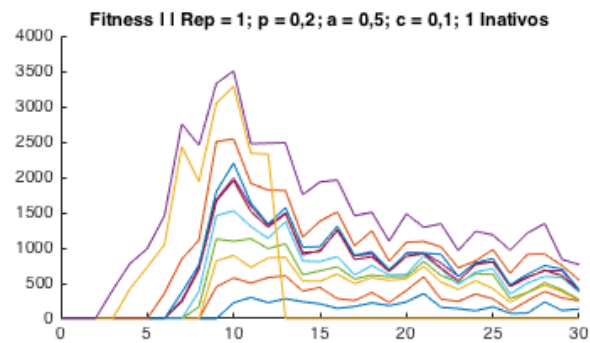
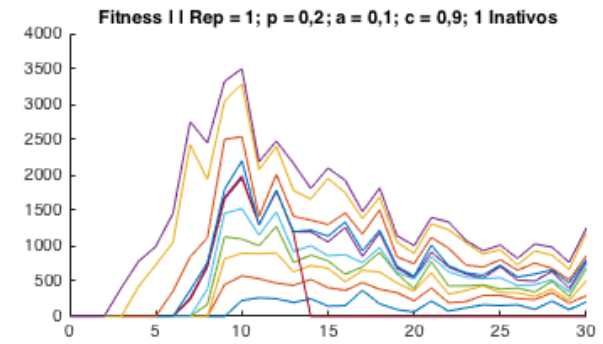
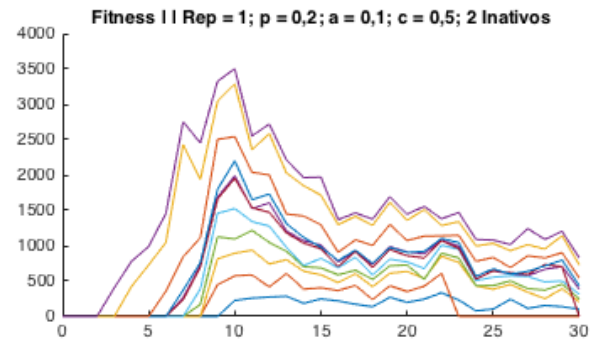
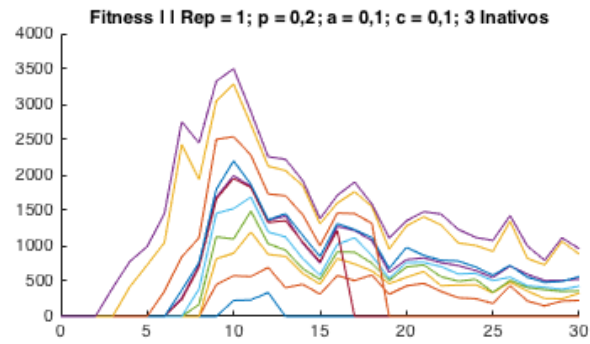
Inativos Estratégicos I | Rep = 1; a = 0,9; c = 0,5



Inativos Estratégicos I | Rep = 1; a = 0,9; c = 0,9



RESULTADOS – Periódicos Estratégicos Inativos



- **Considerações Finais**

- Alguns periódicos tornam-se inativos durante o Comportamento Estratégico.
- O aumento da quantidade de Periódicos Estratégicos não significou necessariamente um aumento da quantidade total de Citações Estratégicas no sistema.
- Maiores níveis de Comportamento Estratégico proporcionam um maior benefício aos Periódicos Estratégicos. Há um maior afastamento relativo em relação aos Periódicos Normais.
- Foi comprovada uma influência positiva do Comportamento Estratégico sobre a centralidade de grau dos Periódicos Estratégicos.
- A reciprocidade periférica foi capaz de afetar a medida de centralidade relacionada à reputação dos periódicos.
- Houve aumento da medida de centralidade relacionada à reputação dos periódicos.

- **Trabalhos Futuros:**

- Análise de Modularidade.
- Análise empírica, comparando a distribuição obtida na simulação com dados reais obtidos na base do Web of Knowledge.